```
Db   617 --------GlnSerLysAsnLysGluThrGlyAspValLeuSerValThrProArgTyr  633
                        |||                         ||||||   :::
Qy  1711 ATGGTGAATCGAAACTGAACTGGCAGATCACCGAGAGGTGGCATCATGGCTGGTGCC  1770
             :::||||:::   ||||:||||     ||||:::::::
Db   634 ThrLeuAsnSerMetLeuAspThrTrpGlnAlaThrAspLeuSerLeuGlnAlaThrVal  653

Qy  1771 CGTTATCGCGGAAAAACACCACGTTCACCCAGAATTATTCGTCACTGAGCGCTGTACAG  1830
             :::   ||||||                                :::
Db   654 ThrTrpTyrGlyLysGln--------------------------------Lys  660

Qy  1831 AAGAAAGTGTATGATGAGAAAGGA----------------------GAATACCTGAAA  1866
             |||                             :::     |||
Db   661 ProLysLysTyrAspTyrHisGlyTyrSerAlaAsnAspGlnLeuSer  680

Qy  1867 GCCTGGACGGTGGTGGATGCGCAGGTCTGTCGTGGAAGATGACGGATGCCCTGACGCTGAAT  1926
             :::   :||||||||||||:::::::::       |||:|:||
Db   681 ProTyrAlaIleAlaGlyLeuGlyGlyTyrThrTyrArgLeuSerLysAsnLeuSerLeuGly  700

Qy  1927 GCTGCGGTGAATAACCTGCTCAACAAGGATTACAGTGACGTGAGCCTGTACAGTGCCGGT  1986
             |||   |||:||||| |:|||||             :::||
Db   701 AlaGlyValAlaAspAsnLeuPheAspLys------------ArgLeuPheArgAlaGly  715

Qy  1987 AAGAGTACCGTGTATGCCCGTGATTACTTCCAGACGGGATCATCAACAACAGGATATGTG  2046
             :::                     |||:::
Db   716 AsnAlaGlnGlyValValGly--------IleAspGlyAlaGlyAlaThrTyrAsn  732

Qy  2047 ATACCTGAGCGAATTACTGGATGTCGTGAACTATCAGTTC  2088
             :::        :::|||
Db   733 GluProGlyArgThrPheTyrThrSerLeuThrAlaSerPhe  746
```

```
RESULT 5
BTUB_ECOLI
ID   BTUB_ECOLI     STANDARD;      PRT;   614 AA.
AC   P06129;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Vitamin B12 receptor precursor.
GN   BTUB OR BFE OR CER OR DCRC OR B3966.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85130824; PubMed=3882670;
RA   Heller K., Kadner R.J.;
RT   "Nucleotide sequence of the gene for the vitamin B12 receptor protein
RT   in the outer membrane of Escherichia coli.";
RL   J. Bacteriol. 161:904-908(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [3]
RP   SEQUENCE OF 1-5 FROM N.A.
RX   MEDLINE=91154132; PubMed=1999392;
RA   Gustafsson C., Lindstroem P.H., Hagervall T.G., Esberg B.,
RA   Bjoerk G.R.;
RT   "The trmA promoter has regulatory features and sequence elements in
RT   common with the rRNA P1 promoter family of Escherichia coli.";
RL   J. Bacteriol. 173:1757-1764(1991).
RN   [4]
RP   SEQUENCE OF 456-614 FROM N.A.
RC   STRAIN=RDD020;
RX   MEDLINE=93106943; PubMed=8093236;
RA   Dougherty T.J., Thanassi J.A., Pucci M.J.;
RT   "The Escherichia coli mutant requiring D-glutamic acid is the result
RT   of mutations in two distinct genetic loci.";
RL   J. Bacteriol. 175:111-116(1993).
RN   [5]
RP   SEQUENCE OF 21-32.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [6]
RP   MUTAGENESIS OF TONB BOX.
RX   MEDLINE=90078094; PubMed=2687240;
RA   Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C.,
RA   Kadner R.J.;
RT   "Point mutations in a conserved region (TonB box) of Escherichia coli
RT   outer membrane protein BtuB affect vitamin B12 transport.";
RL   J. Bacteriol. 171:6526-6533(1989).
CC   -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC       BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M10112; AAA23524.1; -.
DR   EMBL; U00006; AAC43072.1; -.
DR   EMBL; AE000471; AAC76948.1; -.
DR   EMBL; M57568; -; NOT_ANNOTATED_CDS.
DR   EMBL; L14556; AAA23676.1; -.
DR   PIR; A65204; QRECBT.
DR   EcoGene; EG10126; btuB.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Cobalt transport; Transport; TonB box; Signal;
KW   Phage recognition; Receptor; Complete proteome.
FT   SIGNAL        1     20
FT   CHAIN        21    614       VITAMIN B12 RECEPTOR.
FT   SITE         26     33       TONB BOX.
FT   SITE        597    614       TONB C-TERMINAL BOX.
FT   MUTAGEN      28     28       L->P: INACTIVATES UPTAKE.
FT   MUTAGEN      30     30       V->G: INACTIVATES UPTAKE.
FT   CONFLICT    162    162       A -> G (IN REF. 1).
FT   CONFLICT    377    377       A -> R (IN REF. 1).
SQ   SEQUENCE   614 AA;  68407 MW;  AB43CC46A991FF95 CRC64;
```

Alignment Scores:

Pred. No.:	2.33e-30	Length:	614
Score:	528.50	Matches:	187
Percent Similarity:	42.24%	Conservative:	118
Best Local Similarity:	25.90%	Mismatches:	246
Query Match:	14.16%	Indels:	171
DB:	1	Gaps:	30

```
US-10-625-972-4 (1-2091) x BTUB_ECOLI (1-614)

Qy    19 GCTTCCGTAGTCATTCCCTGT------CTCGGATTTTCAGCCAGCAGCAGCATAGCTGCTGCA  72
            ||  |||:||:|:||::||||      :|||||||||:||||:||||::||| ::||||
Db     5 AlaSerLeuLeuThrAlaCysSerValThrAlaPheSerAlaTrpAlaGlnAspThrSer  24

Qy    73 GAGGATGGTGATGATTGTTCGGCATCCGGCTATGAGAAAAAGTGACTAACGGCCAGCCGCC  132
            ||  :::::|||:||||:::|::|||   :||||||:|::||||:|:|||||
Db    25 ProAspThrLeuValValThrAlaAsnArgPheGluGlnProArgSerThrValLeuAla  44
```

```
Qy   133  AGTGTTCTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCACCGATCTGGCCGGAG  192
               :::|||||||:::::|||           |||      ::   |||
Db    45  ProThrThrValValThrArgGlnAspIleAspTrpGlnSerThrSerValAsnAsp     64

Qy   193  GCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGGTACGGGTAAAACCGGAGGGGCTG  252
               ||||||||    ::   ||||||||||           |||
Db    65  ValLeuArgArgLeuProGlyValAlaAspIleThrGlnLysAsnGlySerGlyGlnLeuSer  84

Qy   253  GAAATCAGCATCCCAGGAATGCCAGCGCCAGTTACACGCTGATACTGATTGATGTGTTCGT  312
               |||||                  :::|||
Db    85  SerIlePheIleArgGlyThrAsnAlaSerHisValLeuValLeuIleAspSerPGlyValArg  104

Qy   313  ------CAGGGCGGAAGCAGTGACGTGACTCCCAACGGTTTTCTGCCATG          357
               ||:||||||||||:::::||
Db   105  LeuAsnLeuAlaGlyValSerGlySerArgAlaAspLeuSer---                117

Qy   358  AATACCGGGTTCATCCCCCTGCCGCGCCATTGAGCGTTATCAGGGGGCG            417
               :||:||          ::::::|||:::||||||
Db   118  ------GlnPheProIleAlaLeuValGlnArgValGluTyrIleArgGlyPro      133

Qy   418  ATGTCCACACTGTATGGCTCTGATGCGCCCGTGTGTGAATATCATTACCAGAAAG       477
               ::::|||                  :::|||
Db   134  ArgSerAlaValTyrTyrGlySerAspAlaPheGlyGlyValValAlaSnIleIleThrThrArg  153

Qy   478  AATGCAGACAAATGGCTCTCTTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAA  537
               ||:||||||||:::      :::
Db   154  ------AspGluProGlyThrGlyIleSerGlnSerAlaGly----               163

Qy   538  TGGGTAACAGCAGCCAGTTAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATTCTGTC   597
               |||||:::::||:||||||
Db   164  TrpGlySerAsnSerTyrGlnAsnTyr-------                          172

Qy   598  AGCCTGCAGTACGGCGTAGCACACAACGAGTCCAGGGT-----TCATCGGTCACATCA  651
               ||||||||:::                 |||||
Db   173  --------AspValSerThrGlnInGlnLeuGlnLeuGlyAspLysAspArgValThrLeu  188

Qy   652  CTGAGCGGATACAGCAGGCACGCGTATTCCTTATCCCACGGAGTCACAGAATTATAAATCTT  711
               |||||           ||:||||||
Db   189  LeuGlyIleAspTyrAlaHisThrHis-----                            196

Qy   712  GGTGCACGTCTTGACTGGAAGGCGTCGAGGCAGGATGTGCTCTGGTTTGATATGGGATACC  771
               ||||:||||:||:                       :::
Db   197  ------------------GlyTyrAspValValAlaAlaTyrGlyAsnThrGly      207

Qy   772  ACCCGCAGCGGTTATGATGATAACCGGATGGGGCAACTGGGGGAGTCGTGACGGGGGTATGAC  831
               ||||||:::|||||||||||                  |||
Db   208  ThrGlnAlaGlnThrAspAsn---AspGlyPheLeuSerLysThrLeuTyrGly---        224

Qy   832  CGGAACCCTGCCGTCATCGAGCGAACCAAAAATTTCAGCTGGTATGATCATACTTTCACCTTC  891
               |||||           :::||||:::
Db   225  ------------------AlaLeuGluHisAsnPheThr--                    231

Qy   892  GGAACATGGAAATCGTATCTG----------------AACTGGAACGAG             924
               ::::           |||||||
Db   232  AspAlaTrpSerGlyPheValArgGlyTyrGlyTyrAspAsnArgThrAsnTyrAspAla  251

Qy   925  ACAGAAAATAAAGGTCGTGAGCTTGTACGGCAGT---GTACTGAAGCGGCACAAATGGGGG  981
               ::::           |||:::
Db   252  TyrTyrSerProGlySerProLeuLeuAspThrArgLysLeuTyrSerGlnSerTrpAspAsp  271

Qy   982  CTTGCCGGGTCAGCCGCGGGAGCTAAGGAATCGAAACCTTATCCTGAAATTCATTACTGCTT 1041
               ::||||                  :::
Db   272  ---AlaGly-------LeuArgTyrAsnGlyGluLeuIleLeuLysSerGlnInLeuIle  286

Qy  1042  ACCCCCTCGGGAGAATCT-------------------CATCTGGTTACGGTGTGGGGGGC 1080
               |||:||:||||||
Db   287  ThrSerTyrSerHisSerLysAspTyrAsnTyrAspProHisTyr---------Gly    302

Qy  1081  GAGTTTCAGAGCTCGTCCATCGAAAGACGGAGTT----                     1113
               :::           ::::
Db   303  ArgTyrAspSerSerAlaThrLeuAspGluMetLysGlyLeuTyrThrValGlnInTrpAlaAsn  322

Qy  1114  --GTCCTTGCCAGCACAGGTGAAACTTTCCGGCAGAAAAGCTGG------          1155
```

```
Db   323  AsnValIleuValGlyHisGlySerIleIleGlyAlaAlaGlyValAspTrpGlnLysGlnThrThr  342

Qy  1156  ------TCGGTATTGCTGAGGATGAGTGG                              1185
               :::           ||||||
Db   343  ThrProGlyThrGlyTyrValGluAspGlyTyrAspArgAsnThrGlyIleTyrLeu     362

Qy  1186  ACGGATGCA-------------------CTTGCGCTGACTGCGGGCAGCGCCTATGGACATCAT 1230
               |||                  :::|||
Db   363  ThrGlyLeuGlnInValGlyAspPheThrPheGluGlyIleAlaAlaArgSerAspAspAsn  382

Qy  1231  GAGCAATTCGGGGGACACTTCAGTCCCGTCATATCTGGTCTCGGGATGTGCAGATGCC     1290
               ||||||||           :::
Db   383  SerGlnPheGlyArgHisGlyThrPGlnThrSerAlaGlyTrpGluPheIleGluGlyGly  402

Qy  1291  TGGACGCGTGAAAGGCGGTGTGACCACCGGGATATAAGGCACCCAGAATGGGGCAGCTACAT  1350
               |||||||||           :::|||
Db   403  TyrArgPheIleAlaSerTyrGlyThrGlyThrLysSerAlaProAsnLeuGlyGlnLeuTyr  422

Qy  1351  AAAGGGATTAGTGGTGTGTCCCGGCAGGAGGAAAAACAAATCTACTTGGTAACCCCCGACCTG  1410
               |||||||||           :::
Db   423  -----------------------GlyPheTyrGlyTyrAsnProAsnLeu          430

Qy  1411  AAGCCCGGAAGAGGAGCGGTCAGTTATGAGGCGGGTATTACGATAACCCCGGCCCGGTCTG  1470
               ||:::|||           :::
Db   431  AspProGluLysSerLysGlnInTrpGlu---GlyAlaPheGluGlyLeuThrAlaGlyVal  449

Qy  1471  AATGCCAATGTCACAGGTTTATGACTTCTCCAACAGAGTGTCTCTTATTCCATA         1530
               |||           :::   :::
Db   450  AsnTrpArgIleSerGlyTyrArgAsnAspValSerAsp--LeuIleAspTyr-        466

Qy  1531  AATGATAACACCAATAGCTATGTAAACAGCGGAAAGGCCCGGTTGCACGGTGTGGAATTT  1590
               ::::           :::|||
Db   467  AspAspHisPheArgTyrTyrAsnGluGlyLysLysAlaArgIleLysSerGlyValGluAla  486

Qy  1591  GCCGGCACATGCCGCTGTGGGTCAGAGAGGATGTCACGCTGTCACTGAATTATAACCTGACC  1650
               |||           :::|||
Db   487  ThrAlaAsnPheAspThrGlyProLeuThrHisValSerTyrValSerAspTyrValAspAla  506

Qy  1651  CGAAGTGAACAACGTGATGGTGATAACAAAGGTGCCGCCGCTGAGTTATACCCTGAACAC  1710
               ||:::           :::
Db   507  ArgAsnAlaIleThrAsp----------------ThrProLeuLeuArgArgAlaLysGln  521

Qy  1711  ATGGTGAATGCGAAACTGAACTGCAGATCAACGACATCACCGAAGAGGTGGCATCATGCGGTTGGGGTGCC 1770
               |||:::           ::::||||
Db   522  GlnInValLysTyrGlnInLeuAspTrpGlnInLeuTyrAsp------GlyIle         538

Qy  1771  CGTTATCGCGGAAATGACCACGTCGAGCGTGTAC--AGTGCGCGGTAAGAGTACGCGTGTATGCCGGT 1830
               :::           :::|||||
Db   539  ThrTyrGlnInTyrLeuGlyThrArgTyrAspLysAspTyrSerSer-            553

Qy  1831  AAGAAAGTGTATGATGAGAAAGGAGATAACTTGAAAGCCTGACAGCGGTGGTGGATGCAGGT 1890
               :::           |||||||
Db   554  --------TyrProTyrGlnInThrValLysMetGlyValGlyValSerLeuTrpAspLeuAla 570

Qy  1891  CTTGTCGTGGAAGATGACGGATGCGCCCTACGCGTGAATGCTGCGGTGATAACCTGCTCAAC  1950
               |||||||||           ::::|||
Db   571  ValAlaTyrProValThrSerHisLeuThrValArgGlyLysIleAlaAsnLeuPheAsp     590

Qy  1951  AAGGATTACAGTGCACGTGAGCCTGTAC--AGTGCGCGGTAAGAGTACGCGTGTATGCCGGT  2007
               :::           ::::|||
Db   591  LysAspTyrGluThrValTyrGlnInThrAlaGlyTyrThrLeuSerGly          610

Qy  2008  GATTAC 2013
Db   611  SerTyr 612
```

RESULT 6
BTUB_SALTY

```
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Vitamin B12 receptor precursor.
GN   BTUB OR STM4130.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=93079446; PubMed=1448622;
RA   Wei B.Y., Bradbeer C., Kadner R.J.;
RT   "Conserved structural and regulatory regions in the Salmonella
RT   typhimurium btuB gene for the outer membrane vitamin B12 transport
RT   protein.";
RL   Res. Microbiol. 143:459-466(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC       BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M89481; AAA27031.1; -.
DR   EMBL; AE008893; AAL22968.1; -.
DR   StyGene; SG10030; btuB.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Cobalt transport; Transport; TonB box; Signal;
KW   Receptor; Complete proteome.
FT   SIGNAL        1     20
FT   CHAIN        21    614       VITAMIN B12 RECEPTOR.
FT   SITE         26     33       TONB BOX.
FT   SITE        597    614       TONB C-TERMINAL BOX.
FT   CONFLICT     81     81       G -> R (IN REF. 1).
FT   CONFLICT    134    134       R -> P (IN REF. 1).
FT   CONFLICT    183    183       N -> I (IN REF. 1).
FT   CONFLICT    280    280       R -> S (IN REF. 1).
SQ   SEQUENCE   614 AA;  68525 MW;  9F51F601A6I5FD62 CRC64;
```

Alignment Scores:

Pred. No.:	
Score:	4.63e-29
	510.50
Percent Similarity:	43.06%
Best Local Similarity:	25.61%
Query Match:	13.68%
DB:	1

Length:	614	
Matches:	179	
Conservative:	122	
Mismatches:	253	
Indels:	145	
Gaps:	28	

```
US-10-625-972-4 (1-2091) x BTUB_SALTY (1-614)

Qy   46  TTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTGATGATTGTTCGGCATCCGGCTAT  105
            ||||||||        |||   |||   :::|:|||||||||::||||        :: ::
Db   16  PheSerAlaTrpAlaGlnAspThrSerProAspThrLeuValValThrAlaAsnArgPhe  35
```

```
Qy  106  GAGAAAAGCTGACTAACGCAGCCGCCAGTGTTCTGTGATTAGCCAGGAGGAATTGCAG  165
            :::::::        |||        :::::|:||||:|||::::|||||||:|:||
Db   36  GlnGlnProArgSerAlaValLeuAlaProValThrIleValThrArgGlnAspIleGlu  55
```

```
Qy  166  TCCAGCCAGTACCACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAA  225
            ||                :::         ||||||::|||||||||||||:|||||
Db   56  ArgTrpGlnSerThrSerValAsnAspValLeuArgArgLeuProGlyValAspIleAla  75
```

```
Qy  226  AGTGGTACGGTAAAACCGGAGGGCTGGAAATCAGCATCGAGAATGCCACCGAGTAC  285
            :::||:||||        ||||||||        |||||||||
Db   76  GlnSerGlyGlyAlaGlyGlyGlnAsnSerIlePheIleArgGlyThrAsnSerSerHis  95
```

```
Qy  286  ACGGTCGATACTGATTGATGGTGTTCGT--------CAGGGCGGAACGAGCAGTGAC  330
            ||||||        ||::||||||||||||
Db   96  ValLeuValLeuIleAspGlyValArgLeuAsnLeuAlaGlyValSerGlySerAlaAsp  115
```

```
Qy  331  GTGACTCCCACGGTTTTCTGCCATGAATACCGGGTTCATGCCCCTCTGGCCGCCATT  390
            :::||||||        |||||||
Db  116  GlnPheProValSerLeuVal  124
```

```
Qy  391  GAGCGTATTGAGGTTATCAGGGGGCCGATGTCCACACTGTATGGCCTCTGATCGATGGC  450
            ::|||||||||||||||||||        |||||||||||||||:|
Db  125  GlnArgIleGluTyrIleArgGlyProArgSerAlaIlePheTyrGlySerAspAlaIleGly  144
```

```
Qy  451  GGTGTGGTGAATATCATTACCAGAAAGAATGCAGAAATGGCTCTCTTCCGTCAATGCA  510
            ||||||||||||||||        :::::
Db  145  GlyValValAsnIleIleThrArgArgAsp------------------AsnPro  156
```

```
Qy  511  GGGCTGAATCTGCAGGAAAGCAACAAATGGGTAACGACGCAGCCAGTTTAATTCTGGAGC  570
            ||        |||:|||||:
Db  157  GlyThrGluLeuThrAlaGly------TrpGlySerAsnSerTyrGlnAsnTyr------  172
```

```
Qy  571  AGTGGTCCCCTTGTGGATGATTCTGTCAGCCTGCAGGTACGCGGTAGCACACAACAGCGT  630
            ||||||        ||||||:::
Db  173  -------------------------------------AspIleSerThrGlnGlnGln  179
```

```
Qy  631  CAGGGTTCATCGGTCACATCGGTGAGCGATACCAGCAGGCACGCTATTCCTATCCCACG  690
            :::||||||
Db  180  LeuGly-----------------------GluAsnThrArgAlaThrLeuIle-------  189
```

```
Qy  691  GAGTCACAGAATTATATATCTTGACGTCTTGACTGGAAGGCGTCGGAGCAGGATGTG  750
            |||
Db  190  GlyAspTyrGluTyrThrLysGly-----------------------------------  197
```

```
Qy  751  CTCTGGTTGATATG---------------------GATACCACCCCGCAGCGTTATGATAACCGGGAT  798
            ||||||:::
Db  198  ------FheAspValValAlaAlaLysGlyValGlyThrGlyMetGlnAlaGlnProAlaProArgAsp  215
```

```
Qy  799  GGGCAACTGGGG----AGTCTGACGGGGGGGATAGACCGGACCCTG----CGCTATGAG  849
            ||||        :::
Db  216  GlyPheLeuSerLysThrLeuTyrGlyAlaLeuGluHisThrPheSerAspArgTrpSer  235
```

```
Qy  850  CGAAACAAAATTTCAGCTGGCTATGATCATACATTTCACCTTCGGA-----------  894
            :::
Db  236  GlyPheValArgGlyTyrGlyTyrAspAsnArgThrAspTyrAspAlaTyrTyrSerPro  255
```

```
Qy  895  --------------------------------ACATGGAATCGTATCTGAAC  915
            :::|||
Db  256  GlySerProLeuIleAspThrArgLysLeuTyrSerGlnSerTrpAspAlaGlyLeuHis  275
```

```
Qy  916  TGGAACGAGACAGACAGAAAATAAAGTCGTGAGCTTGTAGCCAGTGTACTG--AAGCCGGAC  972
            :::|||||        ||||||:::|||
Db  276  PheAsn---GlyGluArgIleGlnSerGlnLeuValSerSerTyrSerHisSerLysAsp  294
```

```
Qy  973  ------------------------AAATGGGGCTTGCGGTCAGCCGCGGGAGCTTAAG  1008
            :::::: ::
Db  295  TyrAsnTyrAspProHisTyrGlyArgTyrAspTyrSerAlaThrLeuAspGluMetLys  314
```

```
Qy 1009  GAATCGAGACCTATCCTGAATTCATTACTGCTTACCCCTCTCGGAGAATCTCATCTGTT  1068
            :::||        |||   :::|||||||::||||        :::
Db  315  GlnTyrAsnValGlnTrpThrAsnSerValValValGlyHisGlyAsn---------Val  331
```

```
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000242; AAC74533.1; -.
DR   PIR; F64897; F64897.
DR   EcoGene; EG13774; yncD.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
KW   Complete proteome.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    700       PROBABLE TONB-DEPENDENT RECEPTOR YNCD.
FT   SITE        680    680       TONB C-TERMINAL BOX.
SQ   SEQUENCE   700 AA;  77260 MW;  7B3B9C6ABA48FE4 CRC64;
```

```
Alignment Scores:
Pred. No.:                              3.65e-19       Length:         700
Score:                                  373.50         Matches:        186
Percent Similarity:                     35.04%         Conservative:   109
Best Local Similarity:                  22.09%         Mismatches:     250
Query Match:                            10.01%         Indels:         297
DB:                                     1              Gaps:           37
```

```
US-10-625-972-4 (1-2091) x YNCD_ECOLI (1-700)
```

```
Qy  1069 ACGGTGGGGGGCGAGTTCAGAGCTCGTCATGAAAGACGGAGTTGTCCTGCCAGCACA 1128
                  :::::|||         ::                            --------
Db   332 GlyAlaAlaGlyValAspTrpGlnLysGlnThrThrProGly----------Thr  346

Qy  1129 GGT--------GAAACTTTCCGGCAGAAAAGCTCGTCGTATTGCTGGAGGATGAGTGG 1179
                          |||::::            ||:|||::::
Db   347 GlyTyrValProGluGlyTyrAspGlnArgAsnThrGlyValThrTyrLeuThrGlyLeuGln  366

Qy  1180 CATCTCACGGATGCACTGCCGCCTGACTGCGGGCAGCCGCCTATGAACATCATGAGCAATTC 1239
              |||        |||         ::: |||||       |||||
Db   367 GlnLeuGlyAsp---PheThrLeuGluAlaAlaAlaArgSerAspAspAsnSerGlnPhe  385

Qy  1240 GGGGGACACTTCAGTCCGCGTGCATATCTGGTCTGGATGTGGCAGATGCCTGACGCTG 1299
                  ::|||          |||          :::          ||||
Db   386 GlyArgHisGlyThrArgGlnThrSerGlnArgAlaGlyTrpGluPheIleGluGluGlyTyrArgPhe  405

Qy  1300 AAAGGCGGTGTGACCACGGATATAAGGCACCCAGAATGGGGCAGCTACATAAAGGGATT 1359
                  ::::::|||||||           ::::|||||||
Db   406 IleAlaSerTyrGlyThrSerTyrLysAlaProAsnLeuGlyGlnLeuTyr----------  422

Qy  1360 AGTGTGTGTCCGGCAGGGAAAAACAAATCTACTTGGTAACCCGACTGAAGCCGGAA 1419
                            |||||||:::|||        |||||  |||||
Db   423 --------------GlyTyrTyrGlyAsnProAsnLeuAsnProGlu  433

Qy  1420 GAGAGCGCTCAGTTATGAGGCTGTGGGGTGTATTACGATAACCCCGGTCGAATGCCAAT 1479
                  ::||| |||  |||||     ::| ||
Db   434 LysSerLysGlnTrpGlu---GlyAlaPheGluGlyLeuThrHisAlaGlyValSerTrpArg  452

Qy  1480 GTCACAGGTTTATGACTGACTTCTCCAACAAGATTGTCTTATTCCATAAATGATAAC 1539
          :::::||||:::         |||         ::::|||::::
Db   453 IleSerGlyTyrArgAsnAspIle------AsnAspMetIleAspPheTyr------AspAspHis  469

Qy  1540 ACCAATAGCTATGTAAACAGCGAAGAGGCCCGGTTGCACGGTGTGGAATTGCCGGCACA 1599
                  |||||||        |||||||:::|||        |||||
Db   470 LeuGlnLysTyrTyrTyrAsnGluGlyLysAlaArgIleLysGlyLyIleGluLeuAlaThrAlaAsn  489

Qy  1600 TTGCCGCTGGTGTCAGAGGATGTCACGCTGTCACTGAATTACACCTGGACCCGAAGTGAA 1659
                  ||         ||:::|||||        ::|||||
Db   490 PheAspThrGlyProLeuThrHisThrValSerTyrArgAspTyrValAspTyrValAspArgAsnAla  509

Qy  1660 CAACGTGATCGTGATAACAACAAAGGTCGCGCCCCTGAGGTTATACCCCTGAACACATGGTGAAT 1719
                  |||                        |||
Db   510 IleThrAsp-------------ThrProLeuProArgArgSerLysGlnMetAlaLys  524

Qy  1720 GCGGAAACTGAACTGGCAGATCACCGAAGAGGTGGCATCATGGGCTGGGGTGCCCGGTTATCGC 1779
                  :::::||:::|||||        :::           |||||
Db   525 TyrGlnLeuAspTrpAspValTyrAsp------PheAspTrp---GlyMetThrTyrGln  541

Qy  1780 GGGAAAACACCACGTTCACCAGAATTATTCGTCACTGAGCCGCTGTACAGAGAGAAAGTG 1839
                  :::|||||||                     :::|||||
Db   542 TyrLeuGlySerArgTyrArgAspTyrSerAspTyrSerAla-------------  553

Qy  1840 TATGATGAGAAAGGAGAATACCTGAAAGCCTGGACGGTGGTGATGCAGGTGTCGTGG 1899
             |||                        ::: |||||
Db   554 TyrProTyrArgThrValLysMetGlyGlyValSerLeuGlyValSerLeuThrValAlaTyr  573

Qy  1900 AAGATGACGGATGCCCTGACGCGTGAATGCTGGCGTGAATAACCTGCTCAACAAGGATTAC 1959
                  :::|||||||         ::: |||||
Db   574 ProValThrSerHisLeuThrValArgGlyValLysIleAlaAsnLeuPheAspLysAspTyr  593

Qy  1960 AGTGACGTGAGCCCTGTAC--AGTGCCGGTAAGAAGTACCGCTGTATGCCGTGATTAC 2013
                  :::|||||||        :::|||||
Db   594 GluThrValTyrGlyTyrArgIleTyrGlnThrAlaGlyTyrArgGluTyrThrLeuGlySerTyr  612
```

```
RESULT 7
YNCD_ECOLI
ID   YNCD_ECOLI     STANDARD;      PRT;   700 AA.
AC   P76115;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable tonB-dependent receptor yncD precursor.
GN   YNCD OR B1451.
```

```
Qy     1 ATGCGGAATAACCACTCTGGCTTCCGTAGTCATTCCTGT---CTCCGGATTTCAGCCAGC   57
           ||||||||::|::::|:||||||||:::||:||||:||       :::|||||||
Db     1 MetLysIlePheSerValArgGlnThrArgGlnThrValLeuProAlaLeuLeuValLeuSerProVal  20

Qy    58 AGGATAGTGCTGTCAGAGGATGTGATGATTGTCTCCGGCATCCGGCTATGAGAGAAAAGCTG 117
          |||||||::|||::|||||||||||:||||||:::|||||||||||||||
Db    21 ValPheAlaAlaAspGluGlnValSerAlaIleValSerGluAlaAlaProGlnValValSerGlu  40

Qy   118 ACTAACGCAGCCGCCAGTGTTTCTGTGATTAGCAGCAGGAGAATTGCAG---TCCAGCCAG 174
          :::||::::||        ::|||        |||||||||||
Db    41 LeuAspThrProAlaAlaValSerValValAlaSerValAlaAspGlyLeuMetAlaArgLeuAlaThrPro  60

Qy   175 TACCACGATCTGGCGGAGCCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGTACG 234
          ::|||||::|::|::|||||||::|||         ::|::|
Db    61 ArgIleAsnLeuSerGluSerLeuThrGlyTyrLeuValProGlyLeuValGlnAsnArgGln  80

Qy   235 GGTAAAACCGGAGGGCTGGAAATGCATCCGAGAATGCCAGCCAGTTACACGCTG--- 291
                 ::: |||:::||::|::::||:::||||||||:||||||
Db    81 AsnTyrAlaGlnAspLeuGlnLeuSerIleArgGlyPheGlySerArgSerThrTyrGly 100

Qy   292 --------ATACTGATTGATGGTGTT------------CGTCAGGGGC 318
                  :: :::::        ||::|::
Db   101 IleArgGlyIleArgLeuArgTyrValAspGlyIleIleProAlaThrMetProAspGlyInGlyGly 120
```

```
Qy   319 GGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAATACCGGGTTCATGCCCCT 378
         ::: ||||:::::|
Db   121 GlnThrSerAsnIle-----------------------Asp 126

Qy   379 CTGGCCGCCATTGAGCGTTATTGAGGTTATCAGGGGGCCGATGTCCCACTGTATGCTTCT 438
         |||:::::::::|| ::|||||||||:::|:
Db   127 LeuSerSerValGlnAsnAspHisTyrValThrProGlyAsnGlyAspAspSerGlyAsp--- 145

Qy   439 GATGGCGATGGGCCGGTGTCGTGATAATATCATTACCAGAAGAAATGCAGACAAAATGGCTCTCT 498
         ::||| ||||||:::|:::|
Db   146 AsnAlaSerGlyGlyValMetAsnValThrThrGlnThrGlyGlnGlnPro--------- 162

Qy   499 TCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGGGTAACAGCAGCCAGTT 558
         ::::||||||:::||||:::
Db   163 ------ProThrIleGluGluSerTyrTyrGlySer--------------Phe 174

Qy   559 AATTTCTGGAGCAGTGGTCCCCTTGTGATGATTCTCGTCAGCCTGCAGGTACGCGGTAGC 618
         |||||||
Db   175 GlySerTrpArgTyrGly-----------------------LeuLysAlaThrGlyTyrAla 186

Qy   619 ACACAACAG------CGTCAGGGTCATCGGTCACATCACTGAGCGGATACAGCAGGCACG 672
         ||| :::||
Db   187 ThrGlyAspGlyThrGlnProGlyValAspTyrThrValSerThrArgThrPheThr 206

Qy   673 CGTATTCCTTATCCCACGGAGTCA-----------------CAGAATTATAAATCTTGGT 714
         ||| ::|
Db   207 ThrHisGlyTyrArgAspHisSerArgGlyAlaGlnLysAlaAsnLeuAlaAsnAlaLysLeuGly 226

Qy   715 GCACGTCTTGAC-----------------------------AACAAAATTCAGCTGGC--- 756
         |||::||| ||
Db   227 ValArgIleAspGluAlaSerLysLysSerLeuLeuIlePheAsnSerValAspIleLysLysAla 246

Qy   757 TTTGATATGGATACCACCCGGCAG----------------------CGTTATGATGAT 789
         |||||:::
Db   247 AspAspProGlyGlyLeuThrLysAlaGluThrLysLysProGlnAlaAsnProGlnAlaProArg 266

Qy   790 ----------AACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACC 837
         ::||||||||
Db   267 AlaGluGlnTyrAspValAspThrLysLysArgIleLeuThrGlnThrAlaGlnLeuThrGluTyrGlu 286

Qy   838 CTGCGCCTATGAGCCGA-----------------------------AACAAAATTCAGCTGGC--- 870
         ::: ||||:::
Db   287 ArgSerLeuSerSerArgArgAspMetSerValMetMetTyrArgAlaGlyGluGluThr 306

Qy   871 -----------------TATGATCAT---------------------AAAAACACCACGTTTCACC 879
         |||||:::
Db   307 ThrGlnIleTyrGlnSerIleProMetAlaProGlnLeuAsnProSerHisAlaGlyGlyValVal 326

Qy   880 -----ACTTTCACCTTCGGAACATGGAAATCGTATCTGAACTGGAACGACAGAAAT 933
         :::||||||
Db   327 IleThrLeuGlnArgHisTyrGlnGlyIleTyrAspSerArgTrpThrHisArgGlyGluLeu 346

Qy   934 AAAGGTCGTGCAGCTTGTACCGCAGTGTTGGGGGCTTGCGAAAATGGGGGCTTGCCCGGTCAG 993
         ||||:::::
Db   347 GlyValProValThrPheThrThrGlyLeuAsnTyrGluAsnMetSerGlu--AsnArg 365

Qy   994 CCGCGGGAGCTTAAGGAATCGAACCTTATCCTGAATTCATTACTGCTTACCCCCTCTGGGA 1053
         |||||||||||:::::
Db   366 LysSerGlyTyrTyrAsnAsnPheArgLeuAsnSerGlyMetProGluTyrGlyLeuLysGlyGlu 385

Qy  1054 GAATTCATCTGGTTACGGTGTGGGGGGGCGAGTTTCAGAGCTCGTCCATGAAAGACGGAGTT 1113
         |||||||
Db   386 IleThrLeuGlnArgHisTyrGlnGlyIleTyrAspSerArgTrpThrHisArgGlyGluLeu 394

Qy  1114 GTCCTTGCCAGCAGCACAGGTGAAACTTTCCGGCAGAAAAAGCTGGTCGGTA------TTTGCT 1167
         |||:::|||::: |||||
Db   395 LeuArgArg--AspGluArgAsnLeuMet------------------- 394
```

```
Qy  1168 GAGGATGAGTGGCATCTCACGGATGACGACTTGCCTGCTGACTCGCGGGCAGCCGCTAT----- 1221
         ::: ||||| |||:::||||:|||  |||  |||:|||||||||
Db   402 GlnThrGlnTrpGlnLeuSerGluLysLeuSerLeuAspLysAlaGlyValArgTyrSerSer 421

Qy  1222 ----------------GAACATCATGAGCAATTCCGGGA 1245
         ::|||:::
Db   422 ValTrpPheAspSerAsnAspHisTyrValThrProGlyAsnGlyAspAspSerGlyAsp 441

Qy  1246 -------CACTTCAGTCCGGTGCATATCTGTCTGGGATGTGGCAGATGCCTGG 1293
         ||||||
Db   442 AlaSerTyrHisLysTrpLeuProAlaGlySerLeuLysTyrAlaMetThrAspAlaTrp 461

Qy  1294 ACGTGAAAGGCGTGTGACCACGGGATATAAGGCACCCAGAATGGGCAGCTACATAAA 1353
         :::||||| :::
Db   462 AsnIleTyrLeuAlaAlaGlyLysArgGlyPheGluThrProThrIleAsnGluLeuSerTyr 481

Qy  1354 GGGATTAGTGGTGTGTCCGGCAGGGAAAAACAAATCTACTTGGTAACCCGACCTGAAG 1413
         ||||||
Db   482 ArgAlaAspGlyGlnSerGlyMet------------AsnLeuGlyLeuLys 494

Qy  1414 CCGGAAGAGAGGCTCAGTTATGAGGCTGTGGGTGTATTACGAGTAACCCCGCCGGTCGTGAAT 1473
         ||||||
Db   495 ProSerThrAsnAspThrIleGlyLeuGlySerLysThrArgIleGlyAspGlyLeu--- 513

Qy  1474 GCCAATGTCACAGGTTTTATGACTGACTTCCCAACAGATTGTCTCTTATTCCATAAAT 1533
         ::::::: :::::
Db   514 LeuSerLeuAlaLeuPheGlnThrThrAspThrAspGluIleValAlaAspSerSerSer 533

Qy  1534 GATAAACACCAATAGCTATGTAAACAGCGGAAAGGCCCGGTTGCACGGTGTGGAATTGCC 1593
         |||
Db   534 GlyGlyArgThrThrTyrLysAsnAlaGlyLysThrArgArgGlnGlyAlaGlyLeuAla 553

Qy  1594 GGCACATTGCCCGTGTCGTCAGAG-----------GATGTCACGCTGTCACTGAATTAC 1641
         ||| ::::::
Db   554 -----TrpAspGlnArgPheAlaLeuGlyAspPheArgValAsnAlaSerTrp 568

Qy  1642 ACCTGGACCCGAAGTGAACAACGTGAT------------------CGTTATGCGCGG 1668
         |||||
Db   569 ThrTrpLeuAspAlaThrTyrArgSerAsnValCysAsnGluGlnAspCysAsnGlyAsn 588

Qy  1669 ----------GGTGATAACAAAGT--GCGCCGTGAGTTATACCCTGAACAC 1710
         :::
Db   589 ArgMetProGlyIleAlaIleArgAsnMetGlyPheAlaSerIleGlyTyrValProGluAsp 608

Qy  1711 ATGGTGAATGCGAAACTGAACTGGCACAGATCACCGAAGAGGTGCATCATGGCGTGGGTGCC 1770
         :::
Db   609 ------------GlyTrpProTyrAlaGlyThrGlu---Ala 616

Qy  1771 CGTTATGCGCGG-------------AAAAACACCACGTTTCACC 1800
         ||| :::
Db   617 ArgTyrMetGlyAspIleMetAlaAspGluAsnThrAlaLysAlaProSerTyrThr 636

Qy  1801 -------------------CAGAATTATTCGTCACTGAGCGCTGTACAGAAG 1833
         ::: |||:::
Db   637 LeuValGlyLeuPheThrGlyTyrLysTyrHisAsnLeuThr--- 652

Qy  1834 AAAGTGTATGATGAGAAAGGAGAATACCTGAAAGCCTGACACGGTGGTCGATGCGGATCTCTG 1893
         ||||||
Db   652 ---------------------GlyArgTyr 652

Qy  1894 TCGTGGAAGATGACGGATGCCCTGACGCTGCGTGAATGCTGAATAACCTGCTCAACAAG 1953
         ||||:::
Db   653 ---------------------ValAspLeuProGlyAlaArgValAspAsnLeuPheAspLys 665

Qy  1954 GATTACAGTGACGTGAGCCTGTACAGTCCGGTAAGAGTAGTACCGTGATTAC 2013
         ::||||
Db   666 GluTyr-----ValGlySerValIleValAsnGluSerAsn--------GlyArgTyr 680

Qy  2014 TTCCAGACGGGATCATCAACAACAGGATATGTGATAACCTGAGCGAATTACTGGATGTCG 2073
         ::::::
Db   681 TyrGluProSer-----ProGlyArgAsnTyrArgGlyValGlyGly 692

Qy  2074 CTGAAC 2079
```

```
Db       693 MetAsn 694
             :::|||

RESULT 8
HEMR_YEREN
ID   HEMR_YEREN     STANDARD;      PRT;   687 AA.
AC   P31499;
DT   01-JUL-1993 (Rel. 26, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemin receptor precursor.
GN   HEMR.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=630;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX   MEDLINE=93049186; PubMed=1425573;
RA   Stojiljkovic I., Hantke K.;
RT   "Hemin uptake system of Yersinia enterocolitica: similarities with
RT   other TonB-dependent systems in Gram-negative bacteria.";
RL   EMBO J. 11:4359-4367(1992).
RN   [2]
RP   REVISIONS.
RC   STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA   Stojiljkovic I.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC       UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC       THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68147; CAA48250.1; -.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal;
KW   Receptor.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    687       HEMIN RECEPTOR.
FT   SITE         44     51       TONB BOX.
FT   SITE        670    687       TONB C-TERMINAL BOX.
SQ   SEQUENCE   687 AA;  75226 MW;  69092EF97F5DC08A CRC64;

Alignment Scores:
Pred. No.:
Score:                 7.82e-18          Length:          687
Percent Similarity:    355.00            Matches:         173
Best Local Similarity: 37.96%            Conservative:    106
Query Match:           23.54%            Mismatches:      284
DB:                    9.51%             Indels:          172
                       1                 Gaps:             34

US-10-625-972-4 (1-2091) x HEMR_YEREN (1-687)
```

```
Qy   22 TCCGTAGTCATTCCCTGT------CTTCGGATTTTCAGCCAGCAGCATAGCTGCTGCA  72
                                 ||| :::::
Db   15 SerLeuAlaIleAlaCysThrLeuSerLeuLeuAlaValGlnAlaAlaAspThrSerSerThr  34

Qy   73 GAG------GATGTGATGATTGTCTCGGCATCCGGCTATGAG 108
         |||      :::    |||:|||
Db   35 GlnThrAsnSerLysLysArgIleAlaAspThrMetValValThrAlaThrGlyAsnGlu  54
```

```
Qy  109 AAAAAGTGACTAACGACGCCCAGTGTTTCTGTGATTAGCCAGGAGGAATTGCAGTCC 168
             ||| :::|||
Db   55 ArgSerSerPheGluAlaProMetMetValThrValGluAlaAspThrProThrSer  74

Qy  169 AGCCAGTACCACGATCTGCGGAGGCTCTGAGATCAGTAGACGGGTGTGGATGTTGAAAGT 228
             :::  |||||| :::|||            ||||:::           |||:::
Db   75 GluThrAlaThrSerAlaThrAspMetLeuArgAsnIleProGlyLeuThrVal---Thr  93

Qy  229 GGTACGGGTAAAACCGGAGGGCTGGAAATCAGCATCCGAGAAATGCCAGCCAGTTACACG 288
             |||:::     |||    :::::::::   |||||||
Db   94 GlySerGlyArgValAsnGlyValGlnAspValThrLeuArgGlyTyrGlyLysGlnGlyVal 113

Qy  289 CTGATACTGATTGATGTGTTCCGTCAGGGCGGAAGCAGTGACGTGACTCCCAACGGTTTT 348
             |||    :::  |||  :::::::::|||::::
Db  114 LeuThrLeuValAspGlyIleLeuArgGlnGly----------ThrAspThrGlyHis 128

Qy  349 TCTGCCATGAATACCGGGTTCATGCCCCCTCGCCCGCCCATTGAGCGTATTGAGGTTATC 408
             :::||| :::|||    :::|||:::::::::|||
Db  129 ------LeuAsnSerThrPheLeuAspPro---AlaLeuValLysArgValGluIleVal 145

Qy  409 AGGGGGCCGATGTCCACACTGTATGGCTCTGATGGCGATGGGGCGGTGTGGTGAATATCATT 468
             |||||||   |||       :::||||||   ||||||:::::::::::
Db  146 ArgGlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyValIleSerTyrGlu 165

Qy  469 ACCAGAGAAGAATGCAGACAAAATGGCTCCTTCCGTCAATGCAGGGCTGAAATCGCAGGAA 528
             ||||||       :::||||||::: :::
Db  166 ThrValAspAlaAlaAspLeuLeuLeuProGlyGlnAsnSerGlyTyrArgValTyrSer 185

Qy  529 AGCAACAAATGGGTAACAGCAGCCAGTTAATTCTGGAGCAGTGT----------------- 576
             ::::|| |||::::      :::||||||
Db  186 AlaAlaAlaThrGlyAspHisSer---PheGlyLeuGlyAlaSerAlaPheGlyArgThr 204

Qy  577 CCCCTTGTGGATGATTCTCAGCCTGCAGGTACGC-----GGTAGCACACAACAGCGT 630
             |||     :::|||||| :::|||
Db  205 AspAspValAspGlyIleLeuSerPheGlyThrArgAspIleGlyAsnIleArgGlnSer 224

Qy  631 CAGGGTTCATCGTCACATCACTGAGCGATACAGCAGCAGCGTATTCCTTATCCCACCG 690
             |||            :::
Db  225 AspGlyPheAsnAla--------------------ProAsn 231

Qy  691 GAGTCACAGAATTATAATCTTGGTGCACGTCTTGACTGGGAAGGCGTCGGAGGCAGGATGTG 750
             :::    :::
Db  232 AspGluThrIleSerAsnValLeuAlaLysGlyThrTrpArgIleAspGlnIleGlnSer 251

Qy  751 CCTCTGGTTGATATGGATACCACCCGGCCAGCGTTATGATAACCGGGATGGG------ 801
             |||                :::
Db  252 Leu------------------SerAlaAsnLeuArgTyrTyrAsnAsnSerAlaLeuGluPro 266

Qy  802 -----------------------------CAACTGGGGAGTCTGACGCGGGG 822
             |||
Db  267 LysAsnProGlnThrSerAlaAlaSerSerThrAsnLeuMetThrArgSerArgSerThrIle 286

Qy  823 GGATATGACCGGACCCTGCGCTGACGCTATGAGCGAAACAAAATTCAGCTGGCTATGATCATACT 882
             :::
Db  287 GlnArgAspAlaGlnLeuLysTyrAsnIleLysProLeu----------------AspGlnGlu 302

Qy  883 TTCACCTTCGGAACATGGAAATCGTATCTGAACTGAACTGAAACAGAACAGAAAAATAAAGGTCGT 942
             :::  |||   :::::::|||
Db  303 TrpLeuAsnAlaThrAlaGlnValTyr------TyrSerGluValGluIleAsnAlaArg 320

Qy  943 GAGCTTGTACGGCCAGTGTACTGAAGCCGCAAAATGGGGCTTGCCGTCAGCCGCGGGAG 1002
             ||| :::
Db  321 Pro-----------------------------GlnGlyThrProGluGlu 327

Qy 1003 --------------------CTTAAGGAATGGAACCTTATCCTGAATTCATTACTTGCTTACCCCT 1047
                                     ||| :::
Db  328 GlyArgLysGlnThrThrLysGlyGlyLysLeuGluAsnArgThrArgLeuPheThrAsp 347

Qy 1048 CTGGGAGAATCTCATCTGGTTACGGTGGGGGGGCGAGTTCAGACGTCGTCCATGAAAGAC 1107
             |||||||||:::||| :::
Db  348 SerPheAlaSerHisLeuLeuThrTyrGlyThrGlu-----------AlaTyrLysGln 363
```

```
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemin receptor precursor.
GN   HMUR OR YPO0283 OR YO543.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM6;
RX   MEDLINE=96386041; PubMed=9026634;
RA   Hornung J.M., Jones H.A., Perry R.D.;
RT   "The hmu locus of Yersinia pestis is essential for utilization of
RT   free haemin and haem-protein complexes as iron sources.";
RL   Mol. Microbiol. 20:725-739(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC       UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC       THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U60647; AAC64866.1; -.
DR   EMBL; AJ414141; CAC89146.1; -.
DR   EMBL; AE013655; AAM84131.1; -.
DR   PIR; AG0035; AG0035.
DR   PIR; T12069; T12069.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW   Complete proteome.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    676       HEMIN RECEPTOR.
FT   SITE         44     51       TONB BOX.
FT   SITE        659    676       TONB C-TERMINAL BOX.
SQ   SEQUENCE   676 AA;  74230 MW;  84ED731CB914ACD3 CRC64;
```

Alignment Scores:
Pred. No.:　　　　　676 AA;　　7.95e-17　　Length:　　676

```
Qy   1108  GGAGTTGTCCTTGCCAGCACAGGTGAAACTTCCGGCAG------------AAAAGTCTGGTCG  1158
                  :::    ||||::||| |   ||     :::
Db   364   GluGlnThrProSerGlyAlaThrGluSerPheProGlnAlaAspIleArgPheGlySer  383

Qy   1159  GTATTTGCTGAGGATGAGTGGCATCTCACGGAT---GCACTTGCGCTGACTGCGGGCAGC  1215
                 :::   |||||||              :::::::||  |||||:::
Db   384   GlyTrpLeuGlnAspGlnIleMetGlulleThrLeuArgAspLeuProValSerIleLeuAlaGlyThr  403

Qy   1216  CGCTATGAACATCATGAGCAACTTCGGGGGACACTTCAGTCCGCGTGCATATCTGGTCTGG  1275
                 |||||:::::::::              |||
Db   404   ArgTyrArgAspAsnTyrArgGly------------SerSerGluGlyTyrAlaAspVal  418

Qy   1276  GATGTGGCCAGATGCCTCGACGCTGAAAGGCGGTGTGACCACG------------  1317
                    ||||||    ||||||  |||::
Db   419   Asp--AlaAspLysTrpSerSerArgGlyAlaValSerValThrProThrAspTrpLeu  437

Qy   1318  -----------GGATATAAGGCACCCAGAATGGGGCAGCTACAT--  1350
                    :::|||||||| |||||
Db   438   MetLeuPheGlySerTyrAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsn  457

Qy   1351  -----AAAGGATTAGT---GGTGTGTCCGGCAGGGAAAAACAAATCTA---CTTGGT  1398
                |||||||         ||  |   ||
Db   458   AspSerLysHisPheSerMetAsnIleMetGlyAsnThrLeuThrAsnTyrTrpValPro  477

Qy   1399  AACCCCGACCTGAAGCCGGAGAGAGCCTCAGTTATGAGGCGTCGGGGTGTATTACGATAAC  1458
                |||||::||||||||||    :::   |||
Db   478   AsnProAsnLeuLysProGluThrGlnLeuThrGlnGluTyrGlyPhe------------  493

Qy   1459  CCCGCCGGTCTGAATGCCAATGTCACAGGTTTTATGACTGACTTCTCCAACAAGATTGTC  1518
                  ||||||          ||||||||||||
Db   494   -----GlyLeuArgPheAlaAsnAspLeuMetMetAlaGluAspAspLeuGlnPheLysAla  511

Qy   1519  TCTTATTCCATAAATGATAACACCAATAGCTATGTAAACAGCCGAAAGGCCCGGTTGCAC  1578
                 ||||||||||:::::::::||
Db   512   SerTyr---PheAspThrAsnAlaLysAspTyrIleSerThrGlyValThrMetAspPhe  530

Qy   1579  GGGTGTGGAATTGCCGGC-----------------------------------  1596
                   |||
Db   531   GlyPheGlyProGlyGlyLeuTyrCysLysAsnCysSerThrTyrSerThrAsnIleAsp  550

Qy   1597  ACATTGCCGCTGTGGTCAGAGGATGTCACG------------CTG  1629
                   :::||  ||||||||||||
Db   551   ArgAlaLysIleThrProGlyTrpAspAlaThrMetThrTyrGlnThrGlnTrpPheAsnLeu  570

Qy   1630  TCACTGAATTACACCTGGACCCGAAGTGAACACGTGATGGTCGATAACGAAAAGTGCGCCG  1689
                 |||           |||     :::    |||
Db   571   GlyLeuAlaTyrAsnArgThrArgGlyLysAsnGln--------AsnThrAsnGluTrp  587

Qy   1690  CTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTCGGCCAGATCACCGAAGAG  1749
                ||                :::   ::
Db   588   LeuAspThrIleAsnProAspThrValThrSerThrLeuAspValSerTyrLysGlyGlnPheLysGlyGly  607

Qy   1750  ------GTGGCATCATGGCTGGGTGCCCGTTATCGCGGGAAAACACCAGTTTCACCCAG  1803
                                           PheAlaAsp
Db   608   GlyPheAlaValGlyTyrTrpIleGlyThr----------------------------  619

Qy   1804  AATTATTCGTCACTGAGCGGCTGTACAGAGAGAAAGTGTATGATGGAGAAGGAGAATACCTG  1863
                   :::||||    :::   |||
Db   620   ArgSerSerArgValSerSer---------------------SerGlyThrProGln  631

Qy   1864  AAAGCCTGACGACGGTGGTGGATGCAGGTCTGTCGTGGGAAGATGACGGAT---------GCC  1914
                   :::|||||||      :::||||:::
Db   632   AlaGlyTyrTyrGlyValAsnAspPheTyrValSerTyrLysGlyGlnGluGlnPheLysGlyGly  651

Qy   1915  CTGACGCTGAATGCTGCGTGAATAACCTGCTCAACAAGGATTAC  1959
                  :::|||           :::   |||
Db   652   MetThrThrThrValLeuGlyLysAsnAlaPheAspLysGlySerGlyTyr  666
```

RESULT 9
```
HMUR_YERPE
ID   HMUR_YERPE     STANDARD;     PRT;     676 AA.
AC   Q56989;
DT   01-NOV-1997 (Rel. 35, Created)
```

```
Score:                341.00      Matches:       169
Percent Similarity:   38.04%      Conservative:  103
Best Local Similarity: 23.64%     Mismatches:    267
Query Match:                      Indels:        176
DB:                               Gaps:          31

US-10-625-972-4 (1-2091) x HMUR_YERPE (1-676)
```

```
Db   321 ---------------------------------------ProGlnGlySerAlaGluGluGlyArgGlnThr 332
                                                 |||    |||    :::     ||||||
Qy  1009 GAATCGAACCTTATCCTG-----AATTCATTACTGCTTACCCCTCGGAGAATTCAT 1062
              :::    |||                          :::  |||
Db   333 ThrGluGlyValLysLeuGluAsnArgThrArgLeuPheIleGluSerProAlaSerHis 352

Qy    55 AGCAGCAGATAGCTGCTGCAGCGATGTGATGATTGTCGGCATCCGGCTATGAGAAAAAG 114
            |||||  :::        |||     ||| |||:::|||  |||||||:::|||
Db    37 SerSerLysHisSerThrAspThrMetValValThrAlaThrArgGlyAsnGluArgSer 56
Qy  1063 CTGGTTACGGTGGGGGGGCGGAGTTCAGAGTCGTCCATGAAAGACGGAGTTGTCCTTGCC 1122
             ||||:||||     ||| ||||:::|||      ||| |||
Db   353 LeuLeuThrValGlyGlyGlyGlyGluPheGlnThrValGlnGluGlnThrProGly--- 368

Qy   115 CTGACTAACGGACCAGCCGCCAGTGTTCTCGTGATTAGCCAGGAGAATTGCCAGTCCAGCCAG 174
            ||||  ||| ||| |||      :::      |||       |||:::|||||||
Db    57 SerPheGluAlaProMetMetValThrValIleGluGlyAsnAlaProThrSerGlnThr 76
Qy  1123 AGCACAGGTGAAACTTTCCGGCAG-----------------------AAAAGCTGTCGGTATTGCTGACGAT 1173
            ||||  |||      :::  |||                       ||||||   :::|||
Db   369 GlyAlaThrGluSerPheProGlnAlaLysIleArgPheSerGlyTrpLeuGlnAsp 388
```

```
Db      621 AlaGlyTyrGlyValAlaAsnAspPheTyrValSerTyrLysGlyGlnGluAlaPheLysGly 640

Qy     1915 CTGACGCTGAATGCTCGCGGTGAATAACCTGCTCAACAAGGATTAC 1959
                :::|||      |||::::|      :::    |||:||
Db      641 MetThrThrMetLeuLeuGlyAsnValPheGluLysGluTyr 655


RESULT 10
ID   FYUA_YERPE     STANDARD;      PRT;   673 AA.
AC   P46359;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pesticin receptor precursor (IRPC).
GN   FYUA OR PSN OR YPO1906 OR Y2404.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6/69;
RA   Rakin A., Heesemann J.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM6;
RX   MEDLINE=95204350; PubMed=7896707;
RA   Fetherston J.D., Lillard J.W., Perry R.D.;
RT   "Analysis of the pesticin receptor from Yersinia pestis: role in
RT   iron-deficient growth and possible regulation by its siderophore.";
RL   J. Bacteriol. 177:1824-1833(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6/69;
RA   Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,
RA   Kunst F., Carniel E., Glaser P.;
RT   "DNA sequence of the 102 kbases unstable region of Yersinia pestis.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC       SIDEROPHORE YERSINIABACTIN.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH
CC       THE FUR PROTEIN.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z35104; CAA84487.1; -.
DR   EMBL; U09530; AAA69906.1; -.
DR   EMBL; AL031866; CAA21395.1; -.
DR   EMBL; AJ414150; CAC90722.1; -.
DR   EMBL; AE013845; AAM85962.1; -.
DR   PIR; A56148; A56148.
DR   PIR; AF0232; AF0232.
DR   HSSP; P05825; 1PEP.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec. 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW   Transport; TonB box; Complete proteome.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    673       PESTICIN RECEPTOR.
FT   SITE         30     37       TONB BOX.
FT   SITE        657    673       TONB C-TERMINAL BOX.
FT   CONFLICT    484    484       G -> D (IN REF. 1).
FT   CONFLICT    514    514       R -> G (IN REF. 1).
SQ   SEQUENCE   673 AA;  73782 MW;  9C39E6010EBCEB2C CRC64;


Alignment Scores:
Pred. No.:              1.02e-16      Length:            673
Score:                  339.50        Matches:           172
Percent Similarity:     39.14%        Conservative:      120
Best Local Similarity:  23.06%        Mismatches:        283
Query Match:            9.10%         Indels:            171
DB:                                   Gaps:               37


US-10-625-972-4 (1-2091) x FYUA_YERPE (1-673)

Qy     1 ATGGCGAATAACC------------------ACTCTGGCTTCCGTAGTCATTCCTGTCTCGGA 45
           ||||::::::|||                  ::::::::::|||       ::::::::
Db     1 MetLysMetThrArgLeuTyrProLeuAlaLeuGlyGlyLeuLeuProAlaIleAla 20

Qy    46 TTTTCAGCCAGCAGCATAGCTGCTCCAGAGGATGTGAATGATTGTCTCGGCATCCGGCTAT 105
           ::::::::|||       ::::::::||||:::::|||||:::::
Db    21 --AsnAlaGlnThrSerGlnGlnAspGluSerThrLeuValThrAlaSerLysGln 39

Qy   106 GAGAGAAAAGCTGACTAACGCA---GCCGCCAGTGTTCTGATTAGCCAGGAGGAATTG 162
           :::    :::|||       |||:::||||||
Db    40 SerSerArgSerAlaSerAlaAsnAsnValSerSerThrValValSerAlaProGluLeu 59

Qy   163 CAGTCCAGCCAGTACCACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTT 222
           :::    :::|||||:::::||:::||||||||:::::::::::
Db    60 SerAspAlaGlyValThrAlaSerAspLysProArgValLeuProGlyLeuAsnIle 79

Qy   223 GAAAGTGGTACGGGTAAAACCCAGGGCTGGAAATCAGCATCCAGGAATGCCAGCCAGT 282
           |||:||||       :::::||:::::||||:|||
Db    80 GluAsn---SerGlyAsnMetLeuPheSerThrIleSerLeuArgGlyValSerSerAla 98

Qy   283 TAC------------------ACGCTGATACTGATTGATGGTGTTCGTCAGGCGCGGAAGC 324
           :::                  ::::|||||||||||     :::||||||
Db    99 GlnAspPheTyrAsnProAlaValThrLeuTyrValAspGlyValProGln---------- 115

Qy   325 AGTGACGTGACTCCCAACGGTTTTCTGCCATGAATACCGGGTTCATGCCCCTCTGGCC 384
           :::|||||       :::|||||:::::||::::||||||:::||:::||
Db   116 ----------------------------------LeuSerThrAsnThrIleGlnLnAlaLeuThr 125

Qy   385 GCCATTGAGCGTATTGAGGTTATCAGGGGCGATGTCCACACTGTATGCTCTGATGCG 444
           :::::: :::||:::|||||||||||       ||||||||||:::::||||
Db   126 AspValGlnSerValGluLeuArgGlyThrLeuGlyTyrLeuTyrLysSerArgIle 145

Qy   445 ATGGGCGGTGGTGGTAATATCATTACCAGAAAGAATGCAGACAAATGGCTCTTCCGTC 504
           :::::|||       :::|||||:::::||||:|||||:::::
Db   146 GlnGlyGlyIleIleAsnIleValThrArgGlnProAspSerThrProArgGlyTyrIle 165
```

Qy 470 AlaLysProPheValAlaLysSerIleAsnTyrGluLysSerIleAsnTyrGluLeuGlyThrArgGluThr 489
              |||||
Db 1459 CCCGCCGGTCGATGCCAATGCCACAGGTTTATGACTGACTTCTCCAACAAGATTGTC 1518

Qy 490 ---AlaAspValThrLeuGlnAlaAlaThrPheTyrThr--HisThrLysAspMetGln 507
              :::
Db 1519 TCTTATTCCATAAATGATAACACCAATAGCTATGTAAACAGCCGAAGGCCGGTTGCAC 1578

Qy 508 LeuTyrSerGlyProValArgMetGlnThrLeuSerAsnAlaGlyLysAlaAspAlaThr 527
Db 1579 GGTGTGGAAATTTGCCGGCACA--------TTGCCGCTGTGGTCAGAGGATGTCACGCTG 1629

Qy 528 GlyValGlyLeuGlnAlaLysProArgPheAlaProGlyTrpSerTrpAspIle------ 545
              ||||||
Db 1630 TCACTGAATTACACCTGGACCCGAAGTGAACACACGTGATGGTGATAAC------AAAGGT 1683

Qy 546 ------AsnGlyAsnValIleArgSerGluPheThrAsnAspSerGluLeuTyrHisGly 563
Db 1684 GCGCGCGCTGAGTGTATACCCCTGAACACATGCGAAACTGACTGGCAGAGATCACC 1743

Qy 564 AsnArgValProPheValProArgTyrGlyAlaGlySerValAlaSerGlyValValIle-- 582
              :::   ::: |||
Db 1744 GAAGAGGTGGCATCATGGCTGGTGCCCGTATCGGCGGGAAAACACCACCGGTTCACCCAG 1803

Qy 583 ---AspThrArgTyrGlyAlaLeuMetProArgLeuAlaVal 595
Db 1804 AATTATTCGTCACTGAGCCGTGTACAGAAGAAAGTGTATGATGAGAAGGAGAATACCTG 1863

Qy 596 Asn------LeuValGlyProHisTyrPheAspGlyAspAsnGlnLeuArg 610
              :::|||
Db 1864 AAA--GCCTGGACGCGTGGTGGATGGCAGGTCTGTCGTGGAAGATGACGGATGCCCTGACG 1920

Qy 611 GlnGlyThrTyrAlaThrLeuAspSerSerLeuGlyTrpGlnAlaThrGluArgMetAsn 630
              :::   :::||||| ::::|||
Db 1921 CTGAATGCTGCCGGTGAATAACCTGCTCAACAAGGATTACAGTGACGTGAGCCTGTACAGT 1980

Qy 631 IleSerValTyrValAspAsnLeuPheAspAspArgArgTyrArgThrTyrGly--TyrMet 649
              :::::   :::::
Db 1981 GCCGGTAAGAGTACGCTG 1998

Qy 650 AsnGlySerSerAlaVal 655
              || |||
Db   AsnGlySerSerAlaVal 655

RESULT 11
FYUA_YEREN
ID FYUA_YEREN   STANDARD;   PRT;   673 AA.
AC P46360;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pesticin receptor precursor (IRPC) (IPR65).
GN FYUA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RT factor with dual function.";
RL Mol. Microbiol. 13:253-263(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1223-75-1, 8081 / Serotype O:8, YE737, and 1209-79;
RA Rakin A., Heesemann J.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC   SIDEROPHORE YERSINIABACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.

Qy 505 AATGCGACAGGCTGAATCTGCGAAAGCAACAAATGGGTAACAGCAGTTAATTTC 564
              |||||| ::   ::::::|||||
Db 166 GlyGlyValSerSerArgAspSerTyrArg---------SerLysPheAsnLeu 181

Qy 565 TGGAGCAGTGGTCCCCTTGTGATGATTCTGTCAGCGTGACGCGTAGCACACA 624
              ||||||:::|| |||
Db 182 ---SerGlyProIleGlnAspGlyLeuLeuTyrGlySerValThrLeuLeuArgGln 199

Qy 625 CAGGCGTCAGGGTTCATCGGTC-------ACATCACTGAGCGATACACAGCAGGCACGCGT 675
              |||||:::||||||
Db 200 ValAspAspAspMetIleAsnProAlaThrGlySerAspSerAspLeuGlyGlyThrArg 219

Qy 676 -------------------ATTCCTTATCCCACGGAGTCACAGAATTATAAATCTTGT 714
                                 ::: ||| :: ::||
Db 220 AlaSerIleAlaGlyAsnValLysLeuArgLeuAlaProAspAspTrpGlnProGluMetGly 239

Qy 715 -----GCACGTTCTGACTGGAAGGCGGTCGGAGCAGGAT------GTGCTCTGGTTTGAT 762
              ::: |||||| ||| ||||
Db 240 PheAlaAlaSerArgArgGluCysThrArgAlaThrGlnAspAlaTyrValGlyTrpAsnAsp 259

Qy 763 ATG---------------------GATACCACCCGGCAGCGCTTATGATAACCGGGAT 798
              ||| ||| |||
Db 260 IleLeuLysGlyLeuSerIleIleSerAspGlySerProAspProTyrMetArgArgCys 279

Qy 799 GGGCAACTGGGGAGGTCGACGGGGGATATGACCCGACCCGGCTATGGAGCGAAACAA 858
              :::||||:::|| ||| |||
Db 280 ThrAspSerGlnThrLeuSerGlyGlyLysTyr---ThrThrAspAspTrpValPheAsnLeu 298

Qy 859 ATTTCAGCT---------GGCTATGATCATACTTTCCACCTTCGGAACATGGAAATCG 906
              ||| |||::: ||||:::
Db 299 IleSerAlaTrpGlnGlnHisGlnTyrSerArgThrPheProSerGlySer--------- 315

Qy 907 TATCTGAACTGGACGAGACGAGAAAATAAAGGTCGTGAGCTTGTACGCAGTGACTGAAG 966
              ::: |||||| ::: |||
Db 316 ----------------------------LeuIleValAsnMetProGln 322

Qy 967 CGCGACAAATGGGGGGCTTGCGGTCAGCGCGGGAGCTTAAGGAATCGAACCTTATCCTG 1026
              ||||||| ||| :: |||
Db 323 Arg------Trp----------------------------AsnGlnAspVal 328

Qy 1027 AATTCATTACTGCTTACCCCTCTGGGAGAAGTCTTCATCTGGTT---ACGGTGGGGGCGAG 1083
              ||| ::||||||:::||||||
Db 329 GlnGluLeuArgAlaAlaThrLeuGlyAspAlaArgThrValAspSerMetValPheGlyLeu 348

Qy 1084 TTTCAGAGCTCGTCC----------------------ATGAAAGACGGAGTTGTC 1116
              ::::: :::::::
Db 349 TyrArgGlnAsnThrArgGluLysLeuAsnSerAlaTyrAspMetProThrProTyr 368

Qy 1117 CTTGCCAGCACAGGTGAAACTTTCCGGCAGAAAAGTCGGTCGTATTTGCTCGAGGATGAG 1176
              :::||:::||||||
Db 369 LeuSerSerThrGlyThrTyrThr---ThrAlaGluThrAlaAlaTyrSerArgAspLeuThr 387

Qy 1177 TGGCATCTCACGGATGCACTTGCCGATGCACTGCCGGCGACTGCCGGTATGAACATCATGAGCAA 1236
              |||||||||||| :: |||
Db 388 TrpHisLeuThrAspArgPheAspIleGlyValGlyLysArgPheSerHisAspLysSer 407

Qy 1237 ---------------------------------------TTCGGG- 1242
                                                     |||||
Db 408 SerThrGlnTyrHisGlySerMetLeuGlyAsnProPheGlyAspGlnGlnGlyLysSerAsn 427

Qy 1243 -----------------GGACACTTCAGTCCGCGTGCATATCTGGTCTCGGGATGTGGCAGAT 1287
                                ||| ||||||
Db 428 AspAspGlnValLeuGlyGlnLeuGluSer------AlaGlyTyrMet-------LeuThrAsp 443

Qy 1288 GCCTGGACGCGCTGAAAGGCGGTGAAGGCGGTGACCACGGGATATAAGGCCACCCCAGAATGGGGCAGCTA 1347
              |||||:::||||| |||
Db 444 AspTrpArgValThrTyrThrArgValAlaAlaGlnAsnGlyTyrArgLysPro 457

Qy 1348 CATAAAAGGGGATTAGTAGTGGTGTGTCCCGGGCAGGAGAAAAACAAATCTACTTGGTAACCCCGAC 1407
              ::: |||
Db 458 --SerGlyTyrAsnIleValProPheArgAlaGlyLeu-----------------AspAsp 469

Qy 1408 CTGAAGCCG--------GAAGAGAGCTCAGTTATGAGGCTGGGGTGTATTACGATAAC 1458

```
CC    -----------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------
DR    EMBL; Z29675; CAA82773.1; -.
DR    EMBL; Z35485; CAA84620.1; -.
DR    EMBL; Z35486; CAA84621.1; -.
DR    EMBL; Z35487; CAA84622.1; -.
DR    EMBL; Z35496; CAA84630.1; -.
DR    PIR; S60142; S60142.
DR    HSSP; P05825; 1FEP.
DR    InterPro; IPR000531; TonB_boxC.
DR    Pfam; PF00593; TonB_dep_Rec; 1.
DR    PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR    PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW    Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW    Transport; TonB box.
FT    SIGNAL        1     22       POTENTIAL.
FT    CHAIN        23    673       PESTICIN RECEPTOR.
FT    SITE         30     37       TONB BOX.
FT    SITE        657    673       TONB C-TERMINAL BOX.
FT    VARIANT     362    362       N -> D (IN STRAINS 8081, YE737 AND 1209-
FT                                 79).
SQ    SEQUENCE   673 AA;  73682 MW;  1C20E0352FAD4DCC CRC64;

Alignment Scores:
Pred. No.:                      3.26e-16          Length:            673
Score:                          332.50            Matches:           172
Percent Similarity:             39.08%            Conservative:      118
Best Local Similarity:          23.18%            Mismatches:        289
Query Match:                    8.91%             Indels:            163
DB:                             1                 Gaps:               36
```

US-10-625-972-4 (1-2091) x FYUA_YEREN (1-673)

```
Qy      1 ATGGGAATAACC----------ACTCTGGCTTCCGCTAGTCCCTGTCTCGGA   45
                                 :::::::::       ::  ::
Db      1 MetLysMetThrArgLeuTyrProLeuAlaLeuGlyLeuLeuLeuProAlaIleAla   20

Qy     46 TTTTCAGCCAGCAGCATAGCTGCTGCAGAGATGTGATGATTGCTCGGCATCCGGCTAT   105
                  :: ::                 :::  :::::||||:|||||
Db     21 --AsnAlaGlnThrSerGlnGlnAspGluSerThrLeuValThrAlaSerLysGln   39

Qy    106 GAGAAAAAGTCGACTAACGCA---GCCGCCAGTGTTTCTGTGATTAGCCAGGAGGAATTG   162
                  :::||||        ||||:|
Db     40 SerSerArgSerAlaSerAlaAsnValSerSerThrValValSerAlaProGluGlu   59

Qy    163 CAGTCCAGCCAGTACCACGATCTGGCGGAGGCGTCTGAGATCAGTAGGAGGGTGTGGATGTT   222
                 :: ::                ||||||:::|:||::::::::::
Db     60 SerAspAlaGlyValThrAlaSerAspLysProArgValLeuProGlyLeuAsnIle   79

Qy    223 GAAAGTGGTACGGTAAAACCGGTAAAACCAGCAAATCAGCAATGCCAGCCAGT   282
                 ||:: ::            |||||:::  ::::::
Db     80 GluAsn---SerGlyAsnMetLeuPheSerThrLeuSerLeuArgGlyValSerSerAla   98

Qy    283 TAC----------------ACGCTGATACTGATTGATGGTGTTCGTCAGGCGCGGAAGC   324
               :: ::                  :|||||||||||||||
Db     99 GlnAspPheTyrAsnProAlaValThrLeuTyrLeuValAlaSerGlyValProGln--   115

Qy    325 AGTGACGTGACTCCCAACGGTTTTTCTGCCATGAATAACCGGGTTCATGCCCCTGGCC   384
                  :: ::                 ::  :|||||
Db    116 --------LeuSerThrAsnThrIleGlnAlaLeuThr---   125

Qy    385 GCCATTGAGCCTATTGAGGTTATCACGGGGGGCCGATGTCCACACNGTATGGCTCTGATGCG   444
               :::  :::||||||||:|::||                 |||||||
Db    126 AspValGlnSerValGluLeuLeuArgGlyProGlnGlyThrLeuTyrGlyLysSerAla   145

Qy    445 ATGGGTGTGTCCGGCAGGAATATCATTACCAGAAAGAATGCAGACAAAAATCACTTGGTAACCCGACCTGAAGCCGTC   504
               :::      |||
Db    146 GlnGlyGlyIleIleAsnIleValThrArgGlnInProAspSerThrProArgGlyTyrIle   165

Db    505 AATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGGGTTAACAGCAGCCAGTTAATTTC   564
                  :::              |||||:|||
Qy    166 GluGlyGlyValSerSerArgAspSerTyrArg---SerLysPheAsnLeu   181

Db    565 TGGAGCAGTGTCCCCTGTGGATGATTCTTCAGCTGCCAGGTACGGCGGTAGCACACAA   624
                 |||:||                 :::
Qy    182 ------SerGlyProIleGlnAspGlyLeuLeuTyrGlySerValThrLeuLeuArgGln   199

Db    625 CAGCGTCAGGGTTCATCGGTC---ACATCACTGAGCGATACAGCAGGCACGCGT   675
                 |||                   |||
Qy    200 ValAspAspGlyMetIleAsnProAlaThrGlySerAspAspLeuGlyGlyThrArg   219

Qy    676 -------------------ATTCCTTATCCCACCGGAGTCACAGAATATAAATCTTGGT   714
                     :::       |||
Db    220 AlaSerIleGlyAsnValLysLeuArgLeuAlaProAspAspGlnTrpGluMetGly   239

Qy    715 ------GCACTCCTGACTGGAAGGCGTCGGACAGGAT---------GTGCTCTGGTTTGAT   762
                  ::            |||
Db    240 PheAlaAlaAlaSerArgGluCysThrArgAlaTyrAlaThrGlnInAspAlaTyrValGlyTrpAsnAsp   259

Qy    763 GATACCACCGGCAGCGTTATGATGAATAACCGGGAT   798
             ATG---                 :::  :::
Db    260 IleLeuLysGlyArgLysLeuSerLeuSerAspGlySerProAspTyrMetArgArgCys   279

Db    799 GGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCTGCGCTATGAGCGAAACAAA   858
                 |||:::||                 ||| :::
Qy    280 ThrAspSerGlnThrLeuSerGlyLysTyr---ThrThrAspAspTrpValPheAsnLeu   298

Qy    859 ATTTCAGCT---------GGCTATGATCATACTTCACCTTCCGAACACATGGAAATCG   906
               ::::::|||||||
Db    299 IleSerAlaTrpGlnGlnGlnHisTyrSerArgThrPheProSerGlySer------Leu   316

Qy    907 TATCTGAACTGGAACGAGACAGAGAAAATAAAGGTCGTGAGCTTGTACGCAGTGTACTGAAG   966
               :::|||                 ::: :::
Db    317 IleValAsnMetProGlnAsnGlnTrpAsnGlnInAspValGlnLeuLeuArgAlaAlaThrLeu   336

Qy    967 CCGGACAAA----------TGGGGGCTTGCCGGTGCAGCCGGGGAGCTTAAG   1008
                  :::              :::
Db    337 GlyAspAlaArgValAspMetValPheGlyLeuTyrArgGlnAsnThrArgGluLys   356

Db   1009 GAATGCGAACCTATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAATCTCATCTGTT   1068
                ||||||
Qy    357 --------------LeuAsnSer   359

Qy   1069 ACGGTGGGGGGGCGGAGTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCTTGCCAGCACA   1128
               ||| :::
Db    360 ----------AlaTyrAsnMetProThrMetProTyrLeuSerSerThr   372

Db   1129 GGTGAAACTTTCCGCAGAAAAGCTGGTCGGTATTCTGGAGGATGAGCATCCACG   1188
                |||||||||                 :::
Qy    373 GlyTyrThr---ThrAlaGluThrLeuAlaAlaAlaTyrSerAspLeuThrTrpHisLeuThr   391

Qy   1189 GATGCACTTGCGCTGACTGCGGCAGCCGCTATGAACATCATGAGCAA----------   1236
               |||: ???            :::
Db    392 AspArgPheAspIleGlyIleGlyValArgPheSerHisAspLysSerSerThrGlnTyr   411

Qy   1237 -------------------------TTCGGG----------   1242
                                        |||
Db    412 HisGlySerMetLeuGlyAsnProPheGlyLysAsnAspGlnVal   431

Db   1243 ---GGACACTTCAGTCCCGCTGCCATATCCGGTCGGATGTGGCAGATGTGGACGCTG   1299
                |||                 :::
Qy    432 LeuGlyGlnLeuSer---AlaGlyTyrMet------LeuThrAspTrpArgVal   447

Qy   1300 AAAGGCGGTGTGACCGGGATATAAGGCGCTACATAAAGGGATT   1359
               |||||||                 :::
Db    448 TyrThrArgIleAlaGlnInGlyTyrLysPro---------SerGlyTyr   460

Db   1360 AGTGGTGTGTCCGGCAGGGAAAACAAATCACTTGGTAACCCGACCTGAAGCCG--   1416
                :::                 |||
Qy    461 AspValGlnSerValGluLeuLeuArgGlyTyrProGlnGlyTyrLeuTyrGlyLysSerAla   504
```

RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY, AND PUTATIVE FUNCTION.
RX MEDLINE=20267815; PubMed=10806384;
RA Molloy M.P., Herbert B.R., Slade M.B., Rabilloud T., Nouwens A.S.,
RA Williams K.L., Gooley A.A.;
RT "Proteomic analysis of the Escherichia coli outer membrane.";
RL Eur. J. Biochem. 267:2871-2881(2000).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT, THAT PARTICIPATES IN
CC   IRON TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AE000182; AAC73892.1; -.
DR EMBL; D90717; BAA35471.1; -.
DR EMBL; D90718; BAA35476.1; -.
DR PIR; E64817; E64817.
DR EcoGene; EG13317; YbiL.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL   1   33   POTENTIAL.
FT CHAIN   34   760   PROBABLE TONB-DEPENDENT RECEPTOR YBIL.
FT SITE   743   760   TONB C-TERMINAL BOX.
SQ SEQUENCE   760 AA; 81960 MW; 1B66B9C2EF68C7CA CRC64;

Alignment Scores:
Pred. No.:   5.82e-14
Score: 301.50   Length: 760
Percent Similarity: 36.36%   Matches: 172
Best Local Similarity: 21.42%   Conservative: 120
Query Match: 8.08%   Mismatches: 258
DB: 1   Indels: 253
                                             Gaps: 40

US-10-625-972-4 (1-2091) x YBIL_ECOLI (1-760)

Qy   37 TGTCTCGGATTTCA-----GCCAGCAGCATAGCTGCT-----------GCAGAG 75
        ||||:||| ||| ::: ||| ||||:|||| |||||
Db   22 CysIleGlyIleThrProValAlaGlnAlaAlaGluGluGlnThrAlaAsnAlaAsp 41

Qy   76 GATGTGATGATTGTCTCGGCATCC----------------------------- 99
        ||::|| :::::|||||||||||
Db   42 AspThrLeuValValGluAlaSerThrProSerLeuTyrAlaProGlnSerAlaAsp 61

Qy 100 --GGCTATGAGAAAAAGCTGACTAACGCACCGCCAGTGTTTCTGTGATTAGCCAGGAG 156
          ||||||||| ||||||:::||||| ||| ||||||
Db   62 ProLysPheSerArgProValAlaAspThrThrArgThrMetThrValIleSerGlyGln 81

Qy 157 GAATTGCAGTCCAGCCAGTACCACCGATCTCGCGGAGGCTCGAGATCAGTAGAGGGTGTG 216
        :::::| :::||| |||||||::||:||||||||||
Db   82 ValIleLysAspGlnGlyAlaThrAsnLeuThrAspAlaLeuLysAsnValProGlyVal 101

Qy 217 GATGTT------------GAAAGTGGTACGGTAAAACCCGAGGGCTGAAATCAGCATC 264
        |||||: :::|| |||||||

```
Db  102 GlyAlaPhePheAlaGlyGluAsnGlyAsnSerThrGlyAsp---AlaIleTyrMet 120
Qy  265 CGAGGATGCCAGCCAGTTACACGTCATGATACTGATTGAGTGTTCGTCAGGGCGGAAGC 324
        |||||   :::   ||  :::   ||   ||||| |||||:|:|||   |||||
Db  121 ArgGlyAlaAspThrSerAsnSer---IleTyrIleAspGlyIleArgAspIleGlySer 139

Qy  325 AGTGACGTGACTCCCAACCGTTTTCTGCCATGAAATACCGGGTTCATGCCCCCTCTGGCC 384
        :::::||||   ::::::   |||   |||
Db  140 -----ValSerArgAspThrPhe------AsnThr------ 147

Qy  385 GCCATTGAGCGTATTGAGGTTATCAGGGGGCGGCCGATGTCCACACTGTCTGATGCG 444
        ||:||:::||||||   ||||||
Db  148 ------GluGlnValGluValIleGlyProSerGlyThrAspTyrThrAspSerAla 165

Qy  445 ATGGGCGCGGTGTGGGTGGAATATCAATTACCAGAAAG------------AATGCA 483
        ||:::::::||:::::::
Db  166 ProThrGlySerIleAsnMetIleSerLysGlnProArgAsnAspSerGlyIleAspAla 185

Qy  484 GACAAATGGCTCTCTTCCGTC--------------------- 504
        :::   |||   |||
Db  186 SerAlaSerIleGlySerAlaTrpHisSerAlaArgGlyThrLeuAspValAsnGlnValIle 205

Qy  505 ------------------------------AATGCAGGGCTG 516
        :::|||||||
Db  206 GlyAspThrThrAlaValArgLeuAsnValMetGlyGluLysThrHisAspAlaGlyArg 225

Qy  517 AATCTGCAGGAAGCAACAAATGGGT-----AACAGCAGTTAATTCTGGAGC 570
        ::   :::::   :::::||   :::::|||
Db  226 AspLysValLysAsnGluArgTyrGlyValAlaProSerValAlaPheGlyLeuGlyThr 245

Qy  571 AGTGGTCCCCTGTG--------------------GATGATTCT 594
        :::   |||
Db  246 AlaAsnAsnArgLeuTyrLeuAsnTyrLeuHisValThrGlnHisAsnThrProAspGlyGly 265

Qy  595 GTC---AGCTGCAGGTACGGCGGTAGCCACCACACAGCGTCAGGGTCATCGGTCACATCA 651
        |||   :::   ::   |||:||||||
Db  266 IleProThrIleGlyLeuProGlyTyrSerAlaProSerAlaGlyThrLeuGlyThr 285

Qy  652 CTGAGGCGGATACAGCAGCCACGCGTATTCCTTATCCACGAGTCACAGAATTATAATCTT 711
        |||   :::::   ||||||||||
Db  286 HisSerGlyLysValAspThrHisAsnPheTyrGlyThrAspSerAsp----------- 301

Qy  712 GGTGCACGTCTTGACTGGAAGGCGTCGGAGCAGGATGTGCTCTGGTTGATATGGAT--- 768
        ||   :::::   ||   |||:||:
Db  302 ------TyrAspAspSerThrThrAspThrAlaThrMetArgPheGluHisAspIle 318

Qy  769 ------------ACCACCCGGCAGCGTTATGATAACCGGGATGGGCAA 804
        ||||||||||
Db  319 AsnAspAsnThrIleArgAsnThrThrArgTrpSerArgValLysGlnAspTyrLeu 338

Qy  805 CTGGGGAGTCGACGGGGGGATATGACCCGACCCTCGCTGTAGAGCGAAACAAAATTCA 864
        :::   ::::::   |||||||
Db  339 MetThrAlaIleMetGlyGlyValAlaSerAsnIleThrGln------------ 351

Qy  865 GCTGCCTATGATCATCACTTTCACCTTCGGAACATGGAAATCGTATCGGAACGAG 924
        ||||::   :::::   :::
Db  352 -------ProThrSerAspValAsnSerTrp-------ThrTrpSerArg 363

Qy  925 ACAGAAATAAAGGTCGTGACGTTGTACGCAGTGTACTGAAGGCGACAAATGGGGGCTT 984
        :::   |||   ::::::
Db  364 ThrAlaAsnThr---LysAspValSerAsnLysIleLeu------------ 375

Qy  985 GCCGGTCAGCCGCGGGAGCTTAAGGAATCGAACCTTATCCTGAATTCATTACTGCTTACC 1044
        :::   :::::   ||||||
Db  376 -----------ThrAsnGlnThrAsnLeu---ThrSerThrPheTyrThrGly 388

Qy 1045 CCTCTGGGAGAATCCATCCGGTTACGGTGGGGGCGAGTT------------ 1086
        ::|||   |||   |||:
Db  389 SerIleGly------HisAspValSerThrGlyValGluPheThrArgArgGluValThr 406

Qy 1087 ------------------------CAGAGCTCG 1095
        |||||:
Db  407 AsnTyrGlyValAsnProValThrLeuProAlaValAsnIleTyrHisProAlaSerSer 426
```

```
Db 1096 TCCATGAAAGACGGAGTTGTCCTTGCGCCACCAGCAGGTGAAACTTTCCGGCAGAAAAGCTGG 1155
Qy  427 IleHisProGlyGlyLeuThrArgAsnGlyAlaAsnAlaAsnGlyGlnThrAspThrPhe 446
        ||||   :::   :::
Db  447 AlaIleTyrAlaPheAspThrLeuGlnIleThrArgAspPheGluAsnGlyGlyIle 466

Qy 1156 TCGGTATTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCGCTGACTCCGGGCAGC 1215
        :::::||||   |||   :::||
Db  467 ArgLeuAspAsnTyrHisThrGluTyrAspSerAlaThrAlaCysGlyGly-----Ser 484

Qy 1216 CGCTATGAA---CATCATGAGCAATTC------------GGGGACACTTCAGT 1254
        |||   :::
Db  485 GlyArgGlyAlaIleThrCysProThrGlyValAla--------LysGlySerProVal 501

Qy 1255 CCGCGTGCATATCTGTCTCGGATGTGGCAGCAGATGCCTGAGCAGGCGGTGTGACC 1314
        |||||
Db  502 ThrThrValAspThrAlaLysSerGlyAsnLeuMetAsnTrpLysGlyAlaAlaLeuTyr 521

Qy 1315 ACGGGATATAAGGCACCCAGAATGGGCAGCTACTATAAAGGATTAGTGGT--- 1365
        ::|||||:
Db  522 HisLeuThrGluAsnGlyAsnValTyrIleAsnTyrAlaValSerGlnIleProProGly 541

Qy 1366 ------------------------------GTGTCCGGGCAG 1377
        |||||:
Db  542 GlyAsnAsnPheAlaLeuAlaGlnSerGlySerGlyAsnSerAlaAsnArgAspPhe 561

Qy 1378 ------------------------GGAAAAACAAATCTACTTGGTAACCCGACCTG 1410
        |||   :::
Db  562 LysProGlnLysAlaAsnThrSerGlulleGlyThrLysTrpGlnValLeuAspLysArg 581

Qy 1411 AAGCCCGAAGAGAGCGTCAGTTATGAGCGTGGGGTGTATTAC-----GATAACCCC 1461
        |||||||||   :::::   ||||||
Db  582 LeuLeuLeuThrAlaAlaLeu------PheArgThrAspIleGluAsnGluValGluGln 599

Qy 1462 GCCGGTCTGAATGCCAATGCAGGTCACAGTTTTATGACTGACTTCTCCAACAGATTGTCTCT 1521
        ||   :::::   ||||||||||
Db  600 --------AsnAspAsp---GlyThrTyrSerGlnAsnTyrGlyIleGlyArgValGluGly 615

Qy 1522 TATTCCATAAATGATAACACCAATAGCATGTAAACAGCGGAAAGGCCCGGTTGCACCGT 1581
        |||||||   :::
Db  616 TyrGluIleSerValAlaGlyAsnIleThrProAlaTrpGln------ValIle 631

Qy 1582 GTGGAATTT---GCCGGCACATTG---CCGCTGGTGGTCAGGAGGATGTCACGCTGTCA 1632
        ||||   :::::
Db  632 GlyGlyTyrThrGlnIleGlnLysAlaThrIleLysAsnGlyLysAspValAlaGlnAspAspGly 651

Qy 1633 CTGAATTACACCTGGACCCGAAGTGAACAACGTGATGGT----GATAACAAA 1680
        ||||||||||   :::::
Db  652 SerSerSerLeuProTyrThrProGluHisAlaPheThrLeuTrpSerGlnTyrGlnAla 671

Qy 1681 GGTGCCGCCTGAGTTATACCCCGAACACATGTGAAACTGAACTGGCCAGATC 1740
        :::   :::::   ||||||
Db  672 ThrAspAspIleSerValGlyAlaGlyAlaArgTyrIleGlySerMetHisLysGlySer 691

Qy 1741 ACCGAAGAGGTGGCCATCATCGCTGGGTGCCCGTTATCGCGGAAA--------- 1785
        ||||::   :::::
Db  692 AspGlyAlaValGlyThrProAlaPheThrGluThrTyr------- 704

Qy 1786 ------------ACACCACGTTCACCCAGAATTATTCGTCACTGAGCCGTGTACAG 1830
        :::   :::::
Db  705 --------Trp---ValAlaAspAlaLys 710

Qy 1831 AAGAAAGTGTATGATGAGAAAGGAGAATACCTGAAAGCCTGGACGCGTGTGGATGCAGGT 1890
        ||||||||||
Db  711 LeuGlyTyrArgValAsnArgAsnLeuAspPheGlnLeuAsnLeuAsnValTyrAsnLeuPheAsp 730

Qy 1891 CTGTCGTGGAAGATGACGGATGCCCTGACGCGGATGAAGTGCTGGAATAACCTGCTCAAC 1950
        :::||   :::::
Db  711 LeuGlyTyrArgValAsnArgAsnLeuAspPheGlnLeuAsnLeuAsnValTyrAsnLeuPheAsp 730

Qy 1951 AAGGATTAC 1959
Db  731 ThrAspTyr 733
```

```
RESULT 13
FECA_ECOLI
ID   FECA_ECOLI      STANDARD;      PRT;    774 AA.
AC   P13036;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Iron(III) dicitrate transport protein feca precursor.
GN   FECA OR B4291.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 34-40.
RC   STRAIN=B;
RX   MEDLINE=88227855; PubMed=2836368;
RA   Pressler U., Staudenmaier H., Zimmermann L., Braun V.;
RT   "Genetics of the iron dicitrate transport system of Escherichia
RT   coli.";
RL   J. Bacteriol. 170:2716-2724(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   van Hove B., Staudenmaier H., Braun V.;
RT   "Novel two-component transmembrane transcription control: regulation
RT   of iron dicitrate transport in Escherichia coli K-12.";
RL   J. Bacteriol. 172:6749-6758(1990).
RN   [4]
RP   SEQUENCE OF 731-774 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=89213950; PubMed=2651410;
RA   Staudenmaier H., van Hove B., Yaraghi Z., Braun V.;
RT   "Nucleotide sequences of the fecBCDE genes and locations of the
RT   proteins suggest a periplasmic-binding-protein-dependent transport
RT   mechanism for iron(III) dicitrate in Escherichia coli.";
RL   J. Bacteriol. 171:2626-2633(1989).
CC   -!- FUNCTION: FECA IS THE OUTER MEMBRANE RECEPTOR PROTEIN IN THE
CC       IRON(III) DICITRATE TRANSPORT SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC       ACTIVE. REGULATION BY THE IRON LEVEL MEDIATED BY THE FUR PROTEIN
CC       AND INDUCTION BY CITRATE PLUS IRON SUGGEST THAT THE IRON(III)
CC       DICITRATE COMPLEX MUST ENTER THE PERIPLASM WHERE IT BINDS TO A
CC       TRANSMEMBRANE PROTEIN, WHICH REGULATES DIRECTLY OR VIA A FURTHER
CC       INDUCTOR, TRANSCRIPTION OF THE FEC GENES.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M20981; AAA23760.1; -.
DR   EMBL; U14003; AAA97187.1; -.
DR   EMBL; AE000499; AAC77247.1; -.
DR   EMBL; M63115; AAA23768.1; -.
DR   EMBL; M26397; AAA23761.1; -.
DR   PIR; E65242; QRECFA.
DR   PDB; 1KMO; 23-OCT-02.
DR   PDB; 1KMP; 23-OCT-02.
```

```
DR   EcoGene; EG10286; fecA.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW   Complete proteome; 3D-structure.
FT   SIGNAL       1     33
FT   CHAIN       34    774       IRON(III) DICITRATE TRANSPORT PROTEIN
FT                              FECA.
FT   SITE        56     63       TONB BOX.
FT   SITE       757    774       TONB C-TERMINAL BOX.
FT   CONFLICT    16     16       T -> A (IN REF. 1).
FT   CONFLICT   190    190       A -> T (IN REF. 1).
FT   CONFLICT   300    300       T -> M (IN REF. 1 AND 3).
FT   CONFLICT   357    357       W -> R (IN REF. 1).
FT   CONFLICT   444    444       L -> V (IN REF. 1).
FT   CONFLICT   749    749       R -> A (IN REF. 1).
SQ   SEQUENCE   774 AA;  85321 MW;  16B5B510276C3B09 CRC64;
```

```
Alignment Scores:
Pred. No.:                          6.91e-14          Length:          774
Score:                              300.50            Matches:         169
Percent Similarity:                 37.52%            Conservative:    109
Best Local Similarity:              22.81%            Mismatches:      282
Query Match:                        8.05%             Indels:          181
DB:                                 1                 Gaps:            40
```

```
US-10-625-972-4 (1-2091) x FECA_ECOLI (1-774)

Qy    64 GCTGCTGCAGAGGATGTGATGATTGTC--------------TCGGCATCCGGCTATGAGAGAAA 111
         |||||         :::                           ||||
Db   109 AlaProLysGluAspAlaLeuThrValValGlyAspTrpLeuGlyAspAlaArgGluAsn 128

Qy   112 AAGCTGACTAACGCAGCCGCCCAGTGTTCTGATTAGCCAGGAGGAATTGCAGTCCAGC 171
         :::         ||||         :::     ||||||:::
Db   129 AspValPheGluHisGlyAlaGlyAlaArgAspValIleArgArgGluAspPheAlaLysThr 148

Qy   172 CAGTACCACGATCTGGCGCGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTT------GAA 225
         |||||         |||  |||  ||         :::      ||||||
Db   149 GlyAlaThrThrMetArgGluValLeuAsnArgIleProGlyValSerAlaProGluAsn 168

Qy   226 AGTGGTACGGTAAAACCGGA---GGGCTGGAAATCAGCATCCGAGGAATGCCA------ 276
         :::|||||||                  |||     :::
Db   169 AsnGlyThrGlySerHisAspLeuAlaMetAsnPheGlyIleArgArgGluAspPheAla 188

Qy   277 --GCCAGTTACACGCTGATACTGATTGATGGTGTTCGTCAGGGCGGAAGCAGTGACGTG 333
         ||||||   ||| |||  ::||||||||::||||||
Db   189 LeuAlaSerArgSerThrValLeuMetAspGlyIle------ProValProPhe 204

Qy   334 ACTCCCAACGTTTTTCTGCCATGAAATACCGGGTTCATGCCCCT-----CTGGCCGCC 387
         |||||         ||||         :::     ||||||
Db   205 AlaProTyrGlyGlnProGlnLeuSer-----------LeuAlaProValSerLeuGlyAsn 221

Qy   388 ATTGAGCGTATTGAGGTTATCAGGGGGCCGATGTCCACACTGTATGGCTCGATGCGATG 447
         :::::  ||||::::  |||||||||||     :::      :::::
Db   222 MetAspAlaIleAspValValArgGlyGlyValGlyAlaValArgTyrProGlnSerVal 241

Qy   448 GGCGGTGTGGTGAAATCATTACCAGGAAAGAATGCAGACAAAATGGCTCTCTTCCGTCAT 507
         ||||||||||||         ||||||         :::
Db   242 GlyGlyValValAsnPheValThrArgAlaIleProGlnAspPhe-----GlyIleGlu 259

Qy   508 GCAGGGCTGAATCTGCAGGAAAGCAACAAATGGGGTAACAGCAGCCAGTTAATTTCTGG 567
         |||||         ||||  ||||||         ||||||||||||||
Db   260 AlaGlyValGluGluGlyGlnLeuSer----------ProThrSerSerGlnAsnAsn------ 274

Qy   568 AGCAGTGGTCCCCCTTGTGGATGATTCTGTCAGCCTGCAGGTACGCGGTAGCACCAACAG 627
         |||||||||||         :::     :::||||||
Db   275 --------ProLysGluThrHisAsnLeuMetValGlyThrArgAlaAspAsn 289

Qy   628 CGTCAGGGTTCATCGGTACATCACTGAGCGGATACAGCAGGCACGCGGTATTCCTTATCCC 687
         |||::::::||  ||      |||  :::
Db   290 GlyPheGlyThrAlaLeuLeu--TyrSerGlyThrArgGlySer------ 303
```

Qy   1588 TTT------------------GCCGGCACATTGCGCCGGCTGTGGTCAGGAGGATGTCACGCTGTCA 1632
                                  ||||||
Db   621 ThrGlnAlaArgTyrAspLeuThrLeuThrProThrLeuAspAsnValSerIleTyr 640

Qy   1633 CTGAATTACACCTGACCCGAAGTGAACACGTGAT---GGTGATAACAAAGGTGCGCCG 1689
                   :::||| || :::   :::
Db   641 AlaSerTyrAlaTyrValAsnAlaGluIleArgGluLysGlyAspThrTyrGlyAsnLeu 660

Qy   1690 CTGAGTTATACCCCTGAACACATGGTGAAATGCGAAACTGACTGCCAGATCACCGAAGAG 1749
             ::: :::|||:||||:||:|||
Db   661 ValProPheSerProLysHis 667

Qy   1750 GTGGCATCATGGCTGGGTGCCCGTATCGCGGGAAAACACCACGTTTCACCCAGAATTAT 1809
                                       ||| |||
Db   668 ---LysGlyThrLeuGlyValGlyValAspTyrLysProGlyAsnTrpThrPheAsnLeuAsn--- 685

Qy   1810 TCGTCCTCGAGCGCTGTACAGAGAAGAAGTGTATGATGAGGAGAAGAATACCTGAAAGCC 1869
                                        ::: :::   :::|||
Db   686 --------SerAspPheGlnSerSerGlnPheAlaAspAsnAlaAsnThrValLysGlu 702

Qy   1870 TGGACGGTGGTGGATGCAGGT--------------------CTGTCGTGG---------- 1899
                                                         |||
Db   703 SerAlaAspGlySerThrGlyArgIleProGlyPheMetLeuTrpGlyAlaArgValAla 722

Qy   1900 ---------AAGATGACGGATGCCCTGACGCTGAATGCTGCGGTGAATAACCTG 1944
                      ::: ||| |||:|||| :::|||
Db   723 TyrAspPheGlyProGlnMetAlaAsp---LeuAsnLeuAlaPheGlyValLysAsnIle 741

Qy   1945 CTCAACAAGGATTACAGTGACGTGACGCTGTACAGTGCCGGTAAGAGTACGCCTGTATGCC 2004
                   :::|||||| :::   :::|||||
Db   742 PheAspGlnAspTyrPhe---IleArgSerTyrAspAspAsnAsnLysGlyIleTyrAla 760

Qy   2005 GGT 2007
            |||
Db   761 Gly 761

RESULT 14
Y262_HAEIN
ID   Y262_HAEIN   STANDARD;   PRT;   723 AA.
AC   P44600;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable tonB-dependent receptor HI0262 precursor.
GN   HI0262.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION BY MASS SPECTROMETRY.
RX   MEDLINE=20137488; PubMed=10675023;
RA   Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA   Gray C., Fountoulakis M.;
RT   "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL   Electrophoresis 21:411-429(2000).
CC   -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.

Qy   688 ACGGAGTCACAGAATTATAATCTTGGTGCAGGAAG--------GCGTCG 738
                                            |||||||
Db   304 --------------------------------------AspTrpArgGluHisSerAlaThr 311

Qy   739 GAGCAGGAGATGTGCTCTGGTTT----------------------GATATGGAT 768
                                                         :::
Db   312 ArgIleAspAspLeuMetLeuLysSerLysTyrAlaProAspGluValHisThrPheAsn 331

Qy   769 ACCACCCGGCAGCGTTATGATAACCGGGATGGGCAACTGGGGAGTCTG--------- 816
            :::|||:|||:||:|||
Db   332 SerLeuLeuGlnTyrTyrAspGluAlaAspMetProGlyGlyLeuSerArgAlaAsp 351

Qy   817 ----------ACGGGGGATATGACCGGACCCTGCGCTATGAGCGA 852
                      ||||||||
Db   352 TyrAspAlaAspArgTrpGlnSerThrArgProTyrAspArgPhe----TrpGlyArg 369

Qy   853 AACAAAATT--TCAGCTGGCTAT-----------------GATCATACTTTCCACCTTC 891
                      ||| ||||| |||
Db   370 ArgLysLeuAlaSerLeuGlyTyrGlnPheGlnProAspSerGlnHisLysPheAsnIle 389

Qy   892 ---GGAACATGGAAATCGTATCTGAACGACGAACAGACAGAAAATAAAGGTCGTGAGCTT 948
               :::   :::|||||
Db   390 GlnGlyPheTyrThrGlnThrArgSerGlyTyrTyrLeuGlu---GlnGlyLysArgIle 408

Qy   949 GTACGCCAGTGTACTGAACGCGCACAAATGGGGCTTGCCCGTCAGCCGCGGGAGCTTAAG 1008
            ||| ::|||:|||
Db   409 ThrLeuSer------ProArgAsnTyrTrpValArgGlyIleGluProArg--------- 423

Qy   1009 GAATCGACCTTATCCTGAATTCATTACTCCTTACCCCTCTGGGAGAATCT---CATCTG 1065
            ||| ::: :: |||:||||
Db   424 TyrSerGlnIlePheMet---------------------IleGlyProSerAlaHisGlu 436

Qy   1066 GTTACGGTGGGGGGCCAGAGTTTCAGAACTCGTCCGTCCATGAAAGACCGGAGTTGTCCTTGCCAGC 1125
            ||| ::: ||| :::
Db   437 ValGlyValGlyTyrArgTyrLeuAsnGlnSerThrHisGluSerThrHisMetArgTyrTyrThrAla 456

Qy   1126 ACAGGTGAAACTTTCCGCAGAAAAGTCTGGTATTTGTCGTAGGAT--------------- 1173
            ||| |||
Db   457 ThrSerSerGlyGlnLeuProSerGlySerProTyrAspArgAspThrArgSerGly 476

Qy   1174 -----GAGTGGCATCTCACGGATGCACTT-----------------GCCGTGACT 1206
                 :::|||:||:|| |||
Db   477 ThrGluAlaHisAlaTrpTyrLeuAspAspLysIleAspGlyAsnTrpIleThrThr 496

Qy   1207 GCGGGAGCCGCTATGAACAATCATGAGCAATTC---------------GGGGGACAC 1248
                                                            |||||||
Db   497 ProGlyMetArgPheGluHisIleGluSerTyrGlnAsnAsnAlaIleThrGlyThrHis 516

Qy   1249 TTCAGT-------------------CCGCGTGCCATATCTGGTCTGGATGTGGCAGAT 1287
                                      |||
Db   517 GluGluValSerTyrAsnAlaProLeuProAlaLeuAsnValLeuTyrHisLeuThrAsp 536

Qy   1288 GCCTGGACGCCTGAAAAGGCGGTGTGACCACGGATATAAGGCACCCAGAATGGGGCAGCTA 1347
            ::||||| ::: |||:|||
Db   537 SerTrpAsnLeuTyrAlaAsnThrGluGlySerPheGlyThrValGlnTyrSerGlnIle 556

Qy   1348 CATAAAGGGATTAGTGCGTGTCCGGCACAGGGAAAAACAAATCTACTTGGTAACCCCGAC 1407
                                                            :::
Db   557 GlyLysAlaValGlnSer----------------GlyAsn------- 564

Qy   1408 CTGAAGCCGGAAGAGAGCCGTCAGTTATGAGGCGTGGGTGTATTACGATAACCCCGCCGT 1467
            :::::||||||::::: ||||||:::
Db   565 ValGluProGluLysAlaArgThrArgGluLeuGlyThrArgAspAlaTyrAspGlyAla-- 583

Qy   1468 CTGAATGCCAATGCACAGGTTTTATGACTGACTTCTCCAACAGATTGTCTCTTATTCC 1527
                                                      |||||||
Db   584 LeuThrIleAlaGluMetGlyLeuPheLeuIleAsnPheAsnAsnGln----------TyrAsp 600

Qy   1528 ATAAATGATAACACCATTAGTGTATGTAAACATGGATATGAAACGACGGAAAGGCCCCGGTTGCACGGTGTGAA 1587
                                                              ||||||||
Db   601 SerAsnGlnThrAsnAspThrValThrAlaArgGlyLysThrArgHisArgHisGlyLeuGlu 620

```
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U32712; AAC21927.1; -.
DR   PIR; C64058; C64058.
DR   TIGR; HI0262; -.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    723
FT   SITE        706    707       PROBABLE TONB-DEPENDENT RECEPTOR HI0262.
SQ   SEQUENCE    723 AA;  80775 MW;  91EB3AB0FFEA2984 CRC64;

Alignment Scores:
Pred. No.:                                    Length:          723
Score:                  3.01e-13              Matches:         166
Percent Similarity:     291.50                Conservative:    116
Best Local Similarity:  36.15%                Mismatches:      275
Query Match:            21.28%                Indels:          223
DB:                     7.81%                 Gaps:             32
                        1

US-10-625-972-4 (1-2091) x Y262_HAEIN (1-723)
```

```
Qy     46  TTTTCAGCCAGCAGCATAGCTGCTCGCAGAGGATGGATGGATGCTCGGCATCCGGCTAT  105
               ||||||      ||||::||| ||      :::|| ::::||  :||||
Db      3  PheSerLysLeuAlaIleThrThrThrLeuValThrAlaAsnAlaLeuAlaGln   22

Qy    106  GAGAAAAAGCTGACTAACGCACCGCCAGCCCAGTGTTTCTGTGATT----------  147
               :::::||  :::||:||                   ::::||    SerIleAsnValIleAlaThrArgAspProSer   38
Db     23  SerValGluLeuAsp----------

Qy    148  -----------------------------AGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC  180
                                       ||||::||  ::::::||| |||
Db     39  ArgPheAlaTyrThrProGluLysProGlnSerLysAspSerLeuLeuSerLysGlnAlaThr   58

Qy    181  GATCTGGCGGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACGGGTAAA  240
               |||::::|||       ||            ||||::|    :::  :::
Db     59  SerValAlaAspAlaLeuGluAspSerIleGluProAsnValAspValArgGlyVal--Arg   77

Qy    241  ACCGGAGGGTGAAATCAGCATCGAGGAAATGCCAGTTACACGCTGATACTGATT  300
               :::||||||     :::|||||| :::|||  ::: ::||||
Db     78  SerIleAlaGlnLysProAsnIleArgGlyLeuSerAspAsnArgValValGlnValIle   97

Qy    301  GATGGTGTTCGTCAGGGGCGGAAGCAGTGACGTGACGGTTTTTCTGCCATGAAT  360
               ||||::||:|||       |||:||:||
Db     98  AspGlyValArgGln-----AsnPheAspLeuValAlaHisArgGly----  110

Qy    361  ACCGGGTTCATGCCCCCTCTGGCCGCCATTGAGGTTATTCAGGGGGCCGATG  420
               :::             ||||||:|| ::::    :||||||
Db    111  ---SerTyrPheLeuProMetSerLeuIleGluIleLysGluLysGlyProSer  129

Qy    421  TCCACACTGTATGGCTCTGATGCGATGGGCGGTGTTGAATAATCATTACC---------  471
               ||||::||::||:||:|| ::::|||||| ::::|| :::: ::
Db    130  SerSerLeuTrpGlySerGlyAlaLeuGlyValValAlaAlaMetArgThrProAsnAla  149

Qy    472  -----------AGAAAGAATGCAGACAAATGGCTCTCCGTCAATCCAGGGCTG-----  516
                         ||||||  ||||||  ::: :::
Db    150  LeuAspLeuLysAsnAsnAspLysPheGlyValLysIleArgGlnGlyTyrGlnThr  169

Qy    517  ------AATCTCCAGGAA--  528
               :::||||||  ||||||
Db    170  AlaAsnAsnLeuSerGluLysSerValPheAlaAlaAsnAspLysAspPheAspVal  189
```

```
Qy    529  -------------------------AGCAACAAATGG  540
                                       ||||||||
Db    190  LeuIleSerGlyPheTyrAsnAlaAspAsnLeuArgThrGlyLysGlyLysLeu  209

Qy    541  GGTAACAGCAGCCAGTTAATTCTGGAGCAGCAGTGGCCCCTTGTGGATGATTCGTCAGC  600
               ||||:::::::::                              ::: :||
Db    210  AsnAsnThrAlaTyrLysGlnPhe------GlyGlyLeuAla----LysPheGly  224

Qy    601  CTGCAGGTACGGCGGTAGCACACAACAGCCTCAGGGTTCATCGGTCACATCACTGAGGCGAT  660
               ||                       :::     :::::||
Db    225  TrpGlnIleAsnAspAlaAsnArgValGluLeuSerHisArgGluThrArgPheLysGln  244

Qy    661  ACAGCAGGCACGCTATTCCTATACCCATGAGGTCACAGAATTATAATCTTGTGCACGT  720
               :::||                         ||
Db    245  ThrAlaProSerAsnAsnGluValGluAsnGluLeuThrAsnGluGlnIleThrAspGln  264

Qy    721  CTT---GACTGGAAGGGTCGGAGCAGCAGATGTGCTCTGGTTTGATATGGATACCACCCGG  777
               :::             ::: |||:||||
Db    265  IleLeuLysLysPheHisGlyGlnLysAspAspLeu------ProProThrGln  281

Qy    778  CAGCGGTTATGATAACCGGATGGGCAACTGGGGACTCTG  816
               :::|||||||  |||::||||||||
Db    282  ProSerProSerGluPheTyrSerLysValLysThrArgLeuGlySerVal  301

Qy    817  ACGGGGGATATGACCGACCCTGGCTATGAGCAGAAACAAAATTTCAGCTGGCTATGAT  876
               :::|||||||||:::                      :::||
Db    302  SerTyrLeuThrAspGlnIlePro------------------------Asp  311

Qy    877  CATACTTTCACCTTCGGGAACATGG----------------AAATCGTATCGAACTGG-  918
               :::||||||  :::                        ||||||||
Db    312  GlnSerThrValPheAsnTyrTyrLeuThrProAspAsnProTyrLeuAsnThrHisIle  331

Qy    919  -----------------------AACGAGACAGAAAATAAAGGTCGTGAGCTTGTACGCAGTGTACTGAAG  966
                                     ||||:::                :::||
Db    332  AlaLeuTyrAsnAsnLysThrIleGluThrIleGluGluGlnIleArgLysValSerGlyVal---  349

Qy    967  CCGCACAAATGGGGCTTGCCGGTCCAGCCCGCGGAGCTTAAGGAATGGAACCTATCCTG  1026
               :::|| ::: :::::|||                 :::||||||
Db    350  LysAspGlnInThrLysLeuThrPheLeuThrValThrArgGlyIleAsnLeuArgSerSerGluLeu---  368

Qy   1027  AATTCATTACTGCTTACCCCCTCGGGAGAATCTCATCTGGTTACCGTGGGGGGC----------  1080
               :::
Db    369  -------------------------SerHisIleSerPheValTyrArgValAsp  378

Qy   1080  ----------------------------------------  1080
                                                   
Db    379  TyrMetArgAspLysIleArgThrGluArgGlyThrAsnGlySerAspAlaLysPheArg  398

Qy   1081  ------GAGTTTCAGAGCTGCTCGTCATGAAAGACGGAGTTGTCCTTGCCAGCACAGTGAA  1134
               :::    :::::|||                     ||||||
Db    399  AlaAspProTyrAsnAlaAsnSerAsnThrThrGlyValTyrLeuLeuIleAlaHisIlePro  418

Qy   1135  ACTTTCCGGACGAAAAGTCGTCGGTATTGCTGGCGATGAGTGCATCCACCGATGCA  1194
               :::||||||      |||||||
Db    419  LeuPheGlyGluLys-----------------------------------------  423

Qy   1195  CTTGCGCTGACTCGCGGCGCAGCCCGCTATGAACATCATGACGCAATTCGGG------  1242
               ||||||:::::||||||         :::
Db    424  LeuLeuValSerProSerValArgTyrAspHisTyrAspThrSerSerLysThrValLys  443

Qy   1243  ----------GGACACTTCAGTCCGCGTGCATATCGGTCTCGGGATGTGGCAGATGCCTG  1293
               :::       ::: |||||||
Db    444  TyrLysAspAsnHisLeuSerProAlaThrLysLeuThrTrpIleValThrAsnTrpLeu  463

Qy   1294  ACGCTGAAAACGCGTGACCACGGATATAAGGCACCCAGAATGGGGCAG----------  1344
               ||||||          :::
Db    464  AspPheThrAlaLysTyrAsnGluAlaPheArgAlaProSerMetGlnGluArgPheVal  483

Qy   1345  ------CTACATAAGGGATTAGTGGTGTGTCCCGGCAGGAAAAACAAATCACTTCGT  1398
               :::   :::        :::
Db    484  SerGlyAlaHisPheGlyAlaAsnThrLeuGlyLeuAspHisIleAsnArgPheValAla  503
```

```
CC        between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC        the European Bioinformatics Institute. There are no restrictions on its
CC        use by non-profit institutions as long as its content is in no way
CC        modified and this statement is not removed. Usage by and for commercial
CC        entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC        or send an email to license@isb-sib.ch).
CC
DR        EMBL; U09840; AAA87059.1; -.
DR        PIR; A57148; A57148.
DR        InterPro; IPR000531; TonB boxC.
DR        Pfam; PF00593; TonB dep Rec; 1.
DR        PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR        PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW        Outer membrane; Transport; TonB box; Signal; Receptor.
FT        SIGNAL        1     19        OR 21 (POTENTIAL).
FT        CHAIN        20    725        HEME/HEMOPEXIN UTILIZATION PROTEIN C.
SQ        SEQUENCE    725 AA;   80838 MW;  FC7886E020CB5BCE CRC64;

Alignment Scores:
Pred. No.:                              3.28e-13        Length:        725
Score:                                  291.00          Matches:       173
Percent Similarity:                     36.27%          Conservative:  115
Best Local Similarity:                  21.79%          Mismatches:    259
Query Match:                            7.80%           Indels:        247
DB:                                     1               Gaps:          33

US-10-625-972-4 (1-2091) x HXC2_HAEIN (1-725)

Qy    46 TTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGGTGATGATGATTGCTCGGCATCCGGCTAT 105
Db     3 PheSerLysLeuSerLeuAlaIleAlaThrThrLeuValThrAlaAsnAlaLeuAlaGln 22

Qy   106 GAGAAAAGTGCTGACTAACGCGACCAGCCGCCAGTGTTTCTGTGATT------------ 147
Db    23 SerIleAsnValIleAlaThrArgAspProSer 38
         SerValGluLeuAsp

Qy   148 -----------------------------------AGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 180
Db    39 ArgPheAlaTyrThrProGluLysGlnSerLysAspSerLeuLeuSerLysGlnAlaThr 58

Qy   181 GATCTGGCGGGAGCCTCTGAGATCAGTAGGAGGGTGTGAGTGGTACGGGTAAA 240
Db    59 SerValAlaAlaLeuAlaLeuGluLeuGluProAsnValAspIleArgGlySer---Arg 77

Qy   241 ACCGGAGGGCTGGAAATCAGCATCGAGGAATGCCAGGCGACCAGTTACACGCTGATACTGATT 300
Db    78 SerIleAlaGlnLysProAsnIleArgGlyLeuSerAspAsnArgValValGlnValIle 97

Qy   301 GATGGTGTTCGTCAGGGGGAGCAGTGACGTGACTCCCAACCGTTTTTCTGCCATGAAT 360
Db    98 AspGlyValArgGln-----AsnPheAspLeuAlaHisArgGly-- 110

Qy   361 ACCGGGTTCATGCCCCCTCGCGCCCATTGAGCGTATTGAGGTTATCAGGGGGCCGATG 420
Db   111 ---SerTyrPheLeuProMetSerLeuLleGlnGluLleGluLlleGlyProSer 129

Qy   421 TCCACACTGTATGGCTCTGATGCGATGGGCGGTGTGGTGAATATCATTACC------- 471
Db   130 SerSerLeuTrpLysSerGlyAlaLeuGlyValValAlaMetArgThrProAsnAla 149

Qy   472 ------------AGAAAGAATGCAGACAAAATGGCTCTCTTCCGTCAATGCAGGCAGGGCTG-- 516
Db   150 LeuAspLeuLeuLysAsnAsnAspLysPheGlyValLysIleArgGlyTyrGlnThr 169

Qy   517 ------AATCTGCAGGAA 528
Db   170 AlaAsnAsnLeuSerGluArgAspValSerValPheAlaAlaAsnAspLysPheAspVal 189

Qy   529 -------AGCAACAAATGG 540
Db   190 LeuIleSerGlyPheTyrAsnAsnAlaAsnLeuArgThrLysLysGlyAsnLysLeu 209

Qy   541 GGTAACAGCAGCCAGTTAATTTCTGGAGCAGTGGTCCCCTTG---TGGATGATTCGTGTC 597
```

```
Qy  1399 AACCCCCGACCTGAAGCCGGAGGAGACGCGTCAGTTATGAG------GCTGGGGTGTATTAC 1452
Db   504 AsnProAsnLeuArgProGluThrAlaLysAsnLysGluIleThrAlaAsnLeuHisPhe 523

Qy  1453 GATAAC------------CCGCCGGTCTGATGCCAATGCACAGTTTATGACTGAC 1500
Db   524 AspSerLeuProPheLysGlnGlyAspLysPheLysIleGluAlaThrTyrPheArgAsnAsp 543

Qy  1501 TTCTCCAACAAGATT------------GTCTCCTATTCC 1527
Db   544 ValLysAspPheIleAsnLeuLysIlePheAsnAspAlaLysThrSerAlaSerAlaGly 563

Qy  1528 ATAAATGATAACACCAAT------------AGCTATGTAAAACAGCGGA 1563
Db   564 AlaAsnProAsnThrAsnGlyAlaLeuAlaLeuProLysAsnSerGlnAsnTyrGlnAsnIleThr 583

Qy  1564 AAGGCCCGGTTGCACGGGTGTGGAATTTGCCGGCACATTGCCCGCTGTCGTCAGGATGTC 1623
Db   584 AsnAlaArgLeuSerGlyIleGluGluLeuGlnAlaGlnTyrGln------ThrGluArgLeu 601

Qy  1624 ACGCTGTCACTGAATTACACCTGGACCCGAACGTGACAACGTGATGGTGATAACAAAGGT 1683
Db   602 ThrLeuPheThrAsnTyrGlySerThrLysGlyLysAspLysAsp--------SerGly 618

Qy  1684 GCGCCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTGGCAGATCACC 1743
Db   619 GluAlaLeuSerAsnIleAlaAlaSerLysIleGlyValGlyValAsnTyrAlaLeuVal 638

Qy  1744 GAAGAGGTGGCATCATGGCTGGGTGCCCGTTATCGCGGGAAAACCACAGTTTCACCCAG 1803
Db   639 LysAsp------------------------------LysPheThrVal 644

Qy  1804 AATTATTCG---TCACTGAGCGCTGTACAGAAGAAGTGTACGAAGAGGAGAAATAC 1860
Db   645 GlyAlaThrValThrHisTyrAlaAlaGlnArgValProLysAspHisSerValThr 664

Qy  1861 CTGAAAGCTGGACGTGGTGGATGCAGGTCGTCG---------------TGG 1899
Db   665 TyrProSerTyrIleLeuThrAspLeuArgAlaThrTyrAlaProLeuLysGlyGluTrp 684

Qy  1900 AAGATGACGGATGCCCTGACGCTGAATGCTGCGGTGAATAACCTGCTCAACAAGGATTAC 1959
Db   685 Lys------------AsnLeuArgLeuAspPheAlaLeuGluLeuGluAsnLeuPheAspArgLysTyr 701
```

```
RESULT 15
HXC2_HAEIN
ID     HXC2_HAEIN     STANDARD;      PRT;     725 AA.
AC     P45557;
DT     01-NOV-1995 (Rel. 32, Created)
DT     01-NOV-1995 (Rel. 32, Last sequence update)
DT     10-OCT-2003 (Rel. 42, Last annotation update)
DE     Heme/hemopexin utilization protein C precursor.
GN     HXUC.
OS     Haemophilus influenzae.
OC     Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC     Pasteurellaceae; Haemophilus.
OX     NCBI_TaxID=727;
RN     [1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=DL42 / Serotype B;
RX     MEDLINE=95270579; PubMed=7751272;
RA     Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
RT     "A gene cluster involved in the utilization of both free heme and
RT     heme:hemopexin by Haemophilus influenzae type b.";
RL     J. Bacteriol. 177:2644-2653(1995).
CC    -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC        CONCENTRATIONS.
CC    -!- SUBCELLULAR LOCATION: Outer membrane.
CC    -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC    -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
CC        H.INFLUENZAE.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
Db   514  ThrAlaAsnLeuHisPheAspSerLeuPheLysLysGlnGlyAspLysPheIleGluAla  533
Qy  1483  ACAGGTTTATGACTGACTTCTCCAACAAGATTGTCTTATTCCATAAATGAT-----      1536
Db   534  ThrTyrPheArgAsnAspValLysAspPheIleProAsnLeuLysIlePheAsnAspAlaLys  553
Qy  1537  ---AACACCAATAGC--                                            1548
Db   554  ThrAsnThrAsnAlaSerAlaSerAlaLeuGlyAlaAlaGlyAlaAlaAsnProAsnGlyAla  573
Qy  1549  -------TATGTAAACAGCGGAGAAGGCCCGGTTGCACGGTGTGGAA              1587
Db   574  LeuLeuProThrLysSerGlnTyrGlnAsnIleThrAsnAlaArgLeuSerGlyIleIleGlu  593
Qy  1588  TTTGCCGGCACATTGCCGCTGTGGTCAGGAGGATGTCACGCTGTCACTGAATTACACCTGG  1647
Db   594  LeuGlnAlaGlnTyrGln------ThrGluArgLeuThrLeuPheThrAsnTyrGlySer    611
Qy  1648  ACCCGAAGTGAACAACGTGATGGTGATAACAAAAGGTGCGCCGCTGAGTTATACCCCTGAA  1707
Db   612  ThrLysGlyGlyValLysAspLysAsp------SerGlyGluAlaLeuSerAsnIleAlaAla  628
Qy  1708  CACATGGTGAATGCGAAACTGACTGGCAGATCACCGAAGAGGTGGCATCATGGCTGGGT    1767
Db   629  SerLysIleGlyValGlyValAsnTyrAlaLeuValLysAsp-                   642
Qy  1768  GCCCGTTATCGCGGGAAAACACCACGTTCACCAGAATTATTCG---TCACTGAGCGCT    1824
Db   643  ---------LysPheThrValGlyAlaValThrHisTyrAla                    654
Qy  1825  GTACAGAAGAAAGTGTATGATGAGAAAGGAGAATACCTGAAAGCCTGGACGGTGGTGGAT  1884
Db   655  AlaGlnArgArgValProLysAspHisSerValThrTyrProSerValTyrIleLeuThrAsp  674
Qy  1885  GCAGGTCGTCG-                                                 1923
Db   675  LeuArgAlaThrTyrAlaProLeuLysGlyGlyGluTrpLys-------AsnLeuArgLeu   691
Qy  1924  AATGCTGCGGTGAATAACCTGCTCAACAAGGATTAC                         1959
Db   692  AspPheAlaLeuGluAsnLeuPheAspArgLysTyr                          703
```

Search completed: October 13, 2004, 11:49:51
Job time : 66.5 secs

```
Db   210  AsnAsnThrAlaTyrLysGlnPhe-GlyGlyLeuAlaLeuAlaLysPheGlyIleLeuAsnAs  229
Qy   598  AGCCTGCAGGTACGCGGTAGCACAC--                                    622
Db   229  pAlaAsnArgValGluLeuSerHisArgGluThrArgPheLysGlnThrArgAlaProSerAs  249
Qy   623  -AACAGCGCTCAGGGTTCATCCGTACCATCATCAGAGCGGATACAGCACGCGTATTCCT    681
Db   249  nAsnGluValGlnAsnGluLeuThrAsnGluGlnIleThrAspGlnIleArgGluPhe--    268
Qy   682  TATCCCACGGAGTCACAGAATTATATCTCGTGCACGTCTTGACTGGAAGGCGTCGGAG    741
Db   269  -------HisLys--ProAsnAsnGlySerProProLysAlaLysProSerGln       283
Qy   742  CAGGATGTGCTCTGGTTTGATAATGGATACACCCGGCAGCGTTATGATAACCGGGATGGG  801
Db   284  GluGlnPhe--------------TyrSerGlyValLysThr                      292
Qy   802  CAACTGGGGAGTCGTCACGGGGGGATGATGACCGGACCCTGCCGTATGAGCGAAACAAAATT  861
Db   293  ArgPheGlySerValSerThrTyrLeuThrArgAspGlnGlnIlePro---------      306
Qy   862  TCAGCTGGCTATGATCATACTTTCACCTTCGGAACATGG--------AAATCGTAT      909
Db   307  ------AspGlnSerThrValPheAsnTyrTyrLeuThrProAspAsnProTyr         322
Qy   910  CTGAACTGG--------AACGAGACAGAAAATAAAGGTCGTGAGCTTGTA            951
Db   323  LeuAsnThrHisIleAlaLeuTyrAsnAsnLysThrIleGluLysGluGlnArgLysVal   342
Qy   952  CGCAGTGTACTGAAGCGCGACAAAATGGGGGCTTGCCGGTCAGCCGCGGGAGCTTAAGGAA  1011
Db   343  SerGlyVal-----LysAspGlnThrLysThrArgGlyIleLeuAsnLeuArgAsn       360
Qy  1012  TCGAACCTTATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAATCTCATCTGGTTACG  1071
Db   361  SerSerGluLeu------------------SerHisIleSerPhe                 369
Qy  1072  GTGGGGGGC-                                                   1080
Db   370  ValTyrGlyValAspTyrMetArgAspLysIleArgThrGluArgGlyThrArgAsnAsnLys  389
Qy  1081  ---------GAGTTCAGAGCGCTCGTCCATGAAAGACGGAGTTGTCCTT            1119
Db   390  AspAlaGlnPheArgAlaAspProTyrAsnAlaAsnSerAsnThrThrGlyValTyrLeu   409
Qy  1120  GCCAGCACAGGTGAAACTTCCGCGACAGAAAAGTGGTCGGTATTTGCTGAGGATGAGTGG  1179
Db   410  IleAlaHisIleProLeuPheGlyGluLys-----------                     419
Qy  1180  CATCTCACGGATGCCGACTTGCGTCGACTGGGCGGCAGCCGCTATGAACATCATGAGCAATTC  1239
Db   420  ------LeuLeuLeuSerProSerValArgTyrArgTyrAspHisTyrAspThrSer     434
Qy  1240  GGG--------GGACACTTCAGTCCCGGTCGCATATCTGGTCTCGGGAT            1278
Db  -435  SerLysThrValLysTyrLysAspAsnHisLeuSerProAlaThrLysLeuThrTrpIle   454
Qy  1279  GTGGCAGATGCCTGGACGCTGAAAGGCGTGTCACCAGGATATAGGCACCCCAGAATG     1338
Db   455  ValThrAsnTrpLeuAspPheThrAlaLysTyrAsnGluAlaPheArgAlaProSerMet   474
Qy  1339  GGGCAG---------CTACATAAAGGATTAGTGGTGTCCGGCAGGGAAAA            1383
Db   475  GlnGluArgPheValSerGlySerHisPheGlyTyrSer---IleLeuGlyArgAsnGlu   493
Qy  1384  ACAAAT--CTACTTGGTAACCCCGACCTGAAGCCGAAGAGAGCGCTCAGTTATGAG----  1437
Db   494  IleAsnLysPheValAlaAlaAsnProAsnLeuArgProGluThrAlaLysAsnLysGluIle  513
Qy  1438  ---GCTGGGGTGTATTACGATAAC-------                              1482
```